

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
WO 01/90359 A2(51) International Patent Classification⁷: **C12N 15/12**, 15/10, C12Q 1/68, C07K 14/705, 16/28, A01K 67/027, A61K 38/17, 39/395, G01N 33/53, 33/577

(21) International Application Number: PCT/US01/16833

(22) International Filing Date: 22 May 2001 (22.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/206,222	22 May 2000 (22.05.2000)	US
60/207,476	25 May 2000 (25.05.2000)	US
60/208,834	2 June 2000 (02.06.2000)	US
60/208,861	2 June 2000 (02.06.2000)	US
60/209,868	7 June 2000 (07.06.2000)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **THORNTON, Michael** [US/US]; 9 Medway Road,Woodside, CA 94062-2612 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **KALLICK, Deborah, A.** [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A2

WO 01/90359

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.

G-PROTEIN COUPLED RECEPTORS**TECHNICAL FIELD**

This invention relates to nucleic acid and amino acid sequences of G-protein coupled receptors and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

10

BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals. Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus activated, triggers an intracellular biochemical cascade that ends with the activation of an intracellular target molecule, such as a transcription factor. This process of signal transduction regulates all types of cell functions including cell proliferation, differentiation, and gene transcription. The G-protein coupled receptors (GPCRs), encoded by one of the largest families of genes yet identified, play a central role in the transduction of extracellular signals across the plasma membrane. GPCRs have a proven history of being successful therapeutic targets.

20

GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic transmembrane domains which together form a bundle of antiparallel alpha (α) helices. GPCRs range in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10; Coughlin, S.R. (1994) Curr. Opin. Cell Biol. 6:191-197). The amino-terminus of a GPCR is extracellular, is of variable length, and is often glycosylated. The carboxy-terminus is cytoplasmic and generally phosphorylated. Extracellular loops alternate with intracellular loops and link the transmembrane domains. Cysteine disulfide bridges linking the second and third extracellular loops may interact with agonists and antagonists. The most conserved domains of GPCRs are the transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account, in part, for structural and functional features of the receptor. In most cases, the bundle of α helices forms a ligand-binding pocket. The extracellular N-terminal segment, or one or more of the three extracellular loops, may also participate in ligand binding. Ligand binding activates the receptor by inducing a conformational change in intracellular portions of the receptor. In turn, the large, third intracellular loop of the activated receptor interacts with a heterotrimeric guanine nucleotide binding (G) protein complex which mediates further intracellular signaling activities, including the activation of second messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and

the interaction of the activated GPCR with ion channel proteins. (See, e.g., Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6; Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176; Baldwin, J.M. (1994) *Curr. Opin. Cell Biol.* 6:180-190.)

5 GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine, γ -aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid
10 mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropin-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli that have yet to be identified are known as orphan receptors.

15 The diversity of the GPCR family is further increased by alternative splicing. Many GPCR genes contain introns, and there are currently over 30 such receptors for which splice variants have been identified. The largest number of variations are at the protein C-terminus. N-terminal and cytoplasmic loop variants are also frequent, while variants in the extracellular loops or transmembrane domains are less common. Some receptors have more than one site at which variance
20 can occur. The splicing variants appear to be functionally distinct, based upon observed differences in distribution, signaling, coupling, regulation, and ligand binding profiles (Kilpatrick, G.J. et al. (1999) *Trends Pharmacol. Sci.* 20:294-301).

25 GPCRs can be divided into three major subfamilies: the rhodopsin-like, secretin-like, and metabotropic glutamate receptor subfamilies. Members of these GPCR subfamilies share similar functions and the characteristic seven transmembrane structure, but have divergent amino acid sequences. The largest family consists of the rhodopsin-like GPCRs, which transmit diverse extracellular signals including hormones, neurotransmitters, and light. Rhodopsin is a photosensitive GPCR found in animal retinas. In vertebrates, rhodopsin molecules are embedded in membranous stacks found in photoreceptor (rod) cells. Each rhodopsin molecule responds to a photon of light by
30 triggering a decrease in cGMP levels which leads to the closure of plasma membrane sodium channels. In this manner, a visual signal is converted to a neural impulse. Other rhodopsin-like GPCRs are directly involved in responding to neurotransmitters. These GPCRs include the receptors for adrenaline (adrenergic receptors), acetylcholine (muscarinic receptors), adenosine, galanin, and glutamate (N-methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkinstall
35 (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 7-9, 19-22,

32-35, 130-131, 214-216, 221-222; Habert-Ortoli, E. et al. (1994) Proc. Natl. Acad. Sci. USA
91:9780-9783.)

The galanin receptors mediate the activity of the neuroendocrine peptide galanin, which
inhibits secretion of insulin, acetylcholine, serotonin and noradrenaline, and stimulates prolactin and
growth hormone release. Galanin receptors are involved in feeding disorders, pain, depression, and
5 Alzheimer's disease (Kask, K. et al. (1997) Life Sci. 60:1523-1533). Other nervous system
rhodopsin-like GPCRs include a growing family of receptors for lysophosphatidic acid and other
lysophospholipids, which appear to have roles in development and neuropathology (Chun, J. et al.
(1999) Cell Biochem. Biophys. 30:213-242).

10 The largest subfamily of GPCRs, the olfactory receptors, are also members of the rhodopsin-
like GPCR family. These receptors function by transducing odorant signals. Numerous distinct
olfactory receptors are required to distinguish different odors. Each olfactory sensory neuron
expresses only one type of olfactory receptor, and distinct spatial zones of neurons expressing distinct
receptors are found in nasal passages. For example, the RA1c receptor which was isolated from a rat
15 brain library, has been shown to be limited in expression to very distinct regions of the brain and a
defined zone of the olfactory epithelium (Raming, K. et al. (1998) Receptors Channels 6:141-151).
However, the expression of olfactory-like receptors is not confined to olfactory tissues. For example,
three rat genes encoding olfactory-like receptors having typical GPCR characteristics showed
expression patterns not only in taste and olfactory tissue, but also in male reproductive tissue
20 (Thomas, M.B. et al. (1996) Gene 178:1-5).

Members of the secretin-like GPCR subfamily have as their ligands peptide hormones such as
secretin, calcitonin, glucagon, growth hormone-releasing hormone, parathyroid hormone, and
vasoactive intestinal peptide. For example, the secretin receptor responds to secretin, a peptide
hormone that stimulates the secretion of enzymes and ions in the pancreas and small intestine
25 (Watson, supra, pp. 278-283). Secretin receptors are about 450 amino acids in length and are found
in the plasma membrane of gastrointestinal cells. Binding of secretin to its receptor stimulates the
production of cAMP.

Examples of secretin-like GPCRs implicated in inflammation and the immune response
include the EGF module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor
30 proteins. These GPCRs are members of the recently characterized EGF-TM7 receptors subfamily.
These seven transmembrane hormone receptors exist as heterodimers in vivo and contain between
three and seven potential calcium-binding EGF-like motifs. CD97 is predominantly expressed in
leukocytes and is markedly upregulated on activated B and T cells (McKnight, A.J. and S. Gordon
(1998) J. Leukoc. Biol. 63:271-280).

35 The third GPCR subfamily is the metabotropic glutamate receptor family. Glutamate is the

major excitatory neurotransmitter in the central nervous system. The metabotropic glutamate receptors modulate the activity of intracellular effectors, and are involved in long-term potentiation (Watson, *supra*, p.130). The Ca²⁺-sensing receptor, which senses changes in the extracellular concentration of calcium ions, has a large extracellular domain including clusters of acidic amino acids which may be involved in calcium binding. The metabotropic glutamate receptor family also includes pheromone receptors, the GABA_B receptors, and the taste receptors.

Other subfamilies of GPCRs include two groups of chemoreceptor genes found in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*, which are distantly related to the mammalian olfactory receptor genes. The yeast pheromone receptors STE2 and STE3, involved in the response to mating factors on the cell membrane, have their own seven-transmembrane signature, as do the cAMP receptors from the slime mold *Dictyostelium discoideum*, which are thought to regulate the aggregation of individual cells and control the expression of numerous developmentally-regulated genes.

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, *supra*). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) Nature 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: luteinizing hormone (precocious puberty); vasopressin V₂ (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism, hypocalcuria, hypercalcemia); parathyroid hormone (short limbed dwarfism); β₃-adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropin (glucocorticoid deficiency) (Wilson, S. et al. (1998) Br. J. Pharmacol. 125:1387-1392; Stadel, J.M. et al. (1997) Trends Pharmacol. Sci. 18:430-437). GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure, and several cardiovascular disorders (Horn, F. and G. Vriend (1998) J. Mol. Med. 76:464-468).

In addition, within the past 20 years several hundred new drugs have been recognized that are directed towards activating or inhibiting GPCRs. The therapeutic targets of these drugs span a wide range of diseases and disorders, including cardiovascular, gastrointestinal, and central nervous system disorders as well as cancer, osteoporosis and endometriosis (Wilson, *supra*; Stadel, *supra*). For example, the dopamine agonist L-dopa is used to treat Parkinson's disease, while a dopamine antagonist is used to treat schizophrenia and the early stages of Huntington's disease. Agonists and antagonists of adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and

tachycardia; serotonin 5HT1D antagonists are used against migraine; and histamine H1 antagonists are used against allergic and anaphylactic reactions, hay fever, itching, and motion sickness (Horn, *supra*).

Recent research suggests potential future therapeutic uses for GPCRs in the treatment of metabolic disorders including diabetes, obesity, and osteoporosis. For example, mutant V2 vasopressin receptors causing nephrogenic diabetes could be functionally rescued *in vitro* by co-expression of a C-terminal V2 receptor peptide spanning the region containing the mutations. This result suggests a possible novel strategy for disease treatment (Schöneberg, T. et al. (1996) EMBO J. 15:1283-1291). Mutations in melanocortin-4 receptor (MC4R) are implicated in human weight regulation and obesity. As with the vasopressin V2 receptor mutants, these MC4R mutants are defective in trafficking to the plasma membrane (Ho, G. and R.G. MacKenzie (1999) J. Biol. Chem. 274:35816-35822), and thus might be treated with a similar strategy. The type 1 receptor for parathyroid hormone (PTH) is a GPCR that mediates the PTH-dependent regulation of calcium homeostasis in the bloodstream. Study of PTH/receptor interactions may enable the development of novel PTH receptor ligands for the treatment of osteoporosis (Mannstadt, M. et al. (1999) Am. J. Physiol. 277:F665-F675).

The chemokine receptor group of GPCRs have potential therapeutic utility in inflammation and infectious disease. (For review, see Locati, M. and P.M. Murphy (1999) Annu. Rev. Med. 50:425-440.) Chemokines are small polypeptides that act as intracellular signals in the regulation of leukocyte trafficking, hematopoiesis, and angiogenesis. Targeted disruption of various chemokine receptors in mice indicates that these receptors play roles in pathologic inflammation and in autoimmune disorders such as multiple sclerosis. Chemokine receptors are also exploited by infectious agents, including herpesviruses and the human immunodeficiency virus (HIV-1) to facilitate infection. A truncated version of chemokine receptor CCR5, which acts as a coreceptor for infection of T-cells by HIV-1, results in resistance to AIDS, suggesting that CCR5 antagonists could be useful in preventing the development of AIDS.

The discovery of new G-protein coupled receptors and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, G-protein coupled receptors, referred to

collectively as "GCREC" and individually as "GCREC-1," "GCREC-2," "GCREC-3," "GCREC-4," "GCREC-5," "GCREC-6," "GCREC-7," "GCREC-8," "GCREC-9," "GCREC-10," "GCREC-11," "GCREC-12," "GCREC-13," "GCREC-14," "GCREC-15," "GCREC-16," "GCREC-17," "GCREC-18," "GCREC-19," "GCREC-20," "GCREC-21," "GCREC-22," and "GCREC-23." In one aspect, 5 the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ 10 ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-23.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ 15 ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-23. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:24-46.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the 20 invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide. In another alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 35 6

90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a)
5 culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid
10 sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ
15 ID NO:1-23.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID
20 NO:24-46, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous

nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional GCREC,

comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the

absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:24-46, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each 35 polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

5 Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

10 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

Table 8 shows tissue-specific expression of polynucleotides of the invention.

15 DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which 20 will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so 25 forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now 30 described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

35 "GCREC" refers to the amino acid sequences of substantially purified GCREC obtained from

any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GCREC. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

An "allelic variant" is an alternative form of the gene encoding GCREC. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding GCREC include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GCREC or a polypeptide with at least one functional characteristic of GCREC. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GCREC, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GCREC. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GCREC. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GCREC is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence.

Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GCREC. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant.

10 Antibodies that bind GCREC polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin,

15 thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"

refers to the capability of the natural, recombinant, or synthetic GCREC, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

“Complementary” describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 5'-TCA-3'.

A “composition comprising a given polynucleotide sequence” and a “composition comprising a given amino acid sequence” refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.

10 Compositions comprising polynucleotide sequences encoding GCREC or fragments of GCREC may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

15 “Consensus sequence” refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

20 “Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows 25 amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
30	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
35	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu

	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
5	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

15 The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least 20 one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such 25 comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of GCREC or the polynucleotide encoding GCREC which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, 30 a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected 35 from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present

embodiments.

A fragment of SEQ ID NO:24-46 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:24-46, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:24-46 is useful, for 5 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:24-46 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:24-46 and the region of SEQ ID NO:24-46 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-23 is encoded by a fragment of SEQ ID NO:24-46. A fragment 10 of SEQ ID NO:1-23 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-23. For example, a fragment of SEQ ID NO:1-23 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-23. The precise length of a fragment of SEQ ID NO:1-23 and the region of SEQ ID NO:1-23 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended 15 purpose for the fragment.

A “full length” polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

“Homology” refers to sequence similarity or, interchangeably, sequence identity, between 20 two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and 25 therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in 30 Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and “diagonals saved”=4. The “weighted” residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polynucleotide sequences.

35 Alternatively, a suite of commonly used and freely available sequence comparison algorithms

is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

15 *Penalty for mismatch: -2*

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

20 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

30 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

35 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative

substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

15 *Open Gap: 11 and Extension Gap: 1 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

20 Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment 25 length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

30 The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

35 "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.

Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive 5 conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

10 Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and 15 conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, 20 for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular 25 circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid 30 sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

35 The words "insertion" and "addition" refer to changes in an amino acid or nucleotide

sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect 5 cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of GCREC which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of GCREC which is useful in any of the antibody production methods disclosed herein or known in 10 the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

15 The term "modulate" refers to a change in the activity of GCREC. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of GCREC.

20 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding 25 sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition.

30 PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

35 "Post-translational modification" of an GCREC may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of GCREC.

"Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

5 "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous 10 nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

15 Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs 20 can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 25 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from 30 megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection 35 programs may also be obtained from their respective sources and modified to meet the user's specific

needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid.

cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or 5 synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

10 The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

15 A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

20 "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

25 "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well 30 as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor 35 of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with

a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be 5 introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having 10 at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater 15 sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the 20 reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The 25 presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-30 1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

35 THE INVENTION

The invention is based on the discovery of new human G-protein coupled receptors (GCREC), the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections.

5 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is
10 denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte 15 polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

20 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the
25 MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these 30 properties establish that the claimed polypeptides are G-protein coupled receptors. For example, SEQ ID NO:2 is 59% identical to rat taste bud receptor protein (GenBank ID g1256389) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.7e-95, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a seven transmembrane receptor (rhodopsin family) domain as
35 determined by searching for statistically significant matches in the hidden Markov model (HMM)-

based PFAM database of conserved protein family domains. The score is 146.3 and the probability value is 2.2e-45. (See Table 3.) In addition, SEQ ID NO:2 contains G-protein coupled receptor signatures as determined by BLIMPS analysis of the BLOCKS (BL00237) and PRINTS (PR00237) databases, and by ProfileScan analysis of the Prosite database, as well as an olfactory receptor signature (PR00245) as determined by BLIMPS analysis of the PRINTS database. Based on BLAST, BLIMPS, ProfileScan, and HMM-based analyses, SEQ ID NO:2 is an olfactory G-protein coupled receptor. In an alternative example, SEQ ID NO:15 is 85% identical to murine odorant receptor MOR18 (GenBank ID g6178008) as determined by BLAST. (See Table 2.) The BLAST probability score is 4.6e-138. SEQ ID NO:15 also contains a seven transmembrane receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:15 is a G-protein coupled receptor. In alternative examples, SEQ ID NO:16 is 72% identical to a mouse olfactory receptor (GenBank ID g3983392) as determined by BLAST analysis, with a probability score of 2.7e-15; SEQ ID NO:17 is 97% identical to a gorilla olfactory receptor (GenBank ID g7211257), with a probability score of 1.2e-109; and SEQ ID NO:18 is 51% identical to a canine olfactory receptor (GenBank ID g1314663), with a probability score of 4.1e-82. (See Table 2.) SEQ ID NO:17 and SEQ ID NO:18 also contain G-protein coupled receptor domains and signature sequences as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16-18 are G-protein coupled receptors. In an alternative example, SEQ ID NO:19 is 56% identical to mouse odorant receptor S19 (GenBank ID g6532001) as determined by BLAST. (See Table 2.) The BLAST probability score is 1.4e-88. SEQ ID NO:19 also contains a seven transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:19 is a G-protein coupled receptor. SEQ ID NO:1, SEQ ID NO:3-14, and SEQ ID NO:20-23 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-23 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention.

Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:24-46 or that distinguish between SEQ ID NO:24-46 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7669623H1 is the identification number of an Incyte cDNA sequence, and NOSEDIC02 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries. Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2525800) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g7329615_000006_002 is the identification number of a Genscan-predicted coding sequence, with g7329615 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows tissue-specific expression of polynucleotides of the invention. Column 1 lists groups of tissues which were tested by polymerase chain reaction (PCR) for expression of the polynucleotides. The remaining columns indicate whether a particular polynucleotide was expressed

in each tissue group. Detection of a PCR product indicated positive expression, denoted by a “+” sign, while inability to detect a PCR product indicated a lack of expression, denoted by a “-” sign.

The invention also encompasses GCREC variants. A preferred GCREC variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GCREC amino acid sequence, and which contains at least one functional or structural characteristic of GCREC.

The invention also encompasses polynucleotides which encode GCREC. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:24-46, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:24-46, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:24-46 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:24-46. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GCREC.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GCREC, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GCREC or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with

which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GCREC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

5 The invention also encompasses production of DNA sequences which encode GCREC and GCREC derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GCREC or any fragment thereof.

10 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:24-46 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in
15 “Definitions.”

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or
20 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA
25 sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

30 The nucleic acid sequences encoding GCREC may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.)
35 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown

sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GCREC may be cloned in recombinant DNA molecules that direct expression of GCREC, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express GCREC.

The nucleotide sequences of the present invention can be engineered using methods generally

known in the art in order to alter GCREC-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or improve the biological properties of GCREC, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GCREC may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.)

Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of GCREC, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.)

The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active GCREC, the nucleotide sequences encoding GCREC or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which 5 contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GCREC. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of 10 sequences encoding GCREC. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GCREC and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding 15 sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression 20 vectors containing sequences encoding GCREC and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 25 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences 30 encoding GCREC. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, 35 N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992)

McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population.

5 (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GCREC. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GCREC can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GCREC into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of GCREC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCREC may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of GCREC. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of GCREC. Transcription of sequences encoding GCREC may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology

(1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GCREC may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses GCREC in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

10 Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

15 For long term production of recombinant proteins in mammalian systems, stable expression of GCREC in cell lines is preferred. For example, sequences encoding GCREC can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in
20 enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These
25 include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat*
30 confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins
35 (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate

luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is 5 also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GCREC is inserted within a marker gene sequence, transformed cells containing sequences encoding GCREC can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GCREC under the control of a single promoter. Expression of the marker gene in response to induction or selection usually 10 indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GCREC and that express GCREC may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR 15 amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GCREC using either 20 specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing 25 monoclonal antibodies reactive to two non-interfering epitopes on GCREC is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and 30 may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GCREC include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GCREC, or any fragments thereof, may be cloned into a vector 35 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for

ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GCREC may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GCREC may be designed to contain signal sequences which direct secretion of GCREC through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "pro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GCREC may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric GCREC protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GCREC activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GCREC encoding sequence and the heterologous protein sequence, so that GCREC may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled GCREC may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

GCREC of the present invention or fragments thereof may be used to screen for compounds that specifically bind to GCREC. At least one and up to a plurality of test compounds may be screened for specific binding to GCREC. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

10 In one embodiment, the compound thus identified is closely related to the natural ligand of GCREC, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which GCREC binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the 15 compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GCREC, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing GCREC or cell membrane fractions which contain GCREC are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GCREC or 20 the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with GCREC, either in solution or affixed to a solid support, and detecting the binding of GCREC to the compound. 25 Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

GCREC of the present invention or fragments thereof may be used to screen for compounds 30 that modulate the activity of GCREC. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for GCREC activity, wherein GCREC is combined with at least one test compound, and the activity of GCREC in the presence of a test compound is compared with the activity of GCREC in the absence of the test compound. A change in the activity of GCREC in the presence of the test compound is 35 indicative of a compound that modulates the activity of GCREC. Alternatively, a test compound is

combined with an in vitro or cell-free system comprising GCREC under conditions suitable for GCREC activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of GCREC may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

5 In another embodiment, polynucleotides encoding GCREC or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the 10 early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce 15 heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

20 Polynucleotides encoding GCREC may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. 25 (1998) *Science* 282:1145-1147).

Polynucleotides encoding GCREC can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding GCREC is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and 30 the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress GCREC, e.g., by secreting GCREC in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

35 **THERAPEUTICS**

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GCREC and G-protein coupled receptors. In addition, the expression of GCREC is closely associated with nasal polyp tissue. Therefore, GCREC appears to play a role in cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections. In the treatment of disorders associated with increased GCREC expression or activity, it is desirable to decrease the expression or activity of GCREC. In the treatment of disorders associated with decreased GCREC expression or activity, it is desirable to increase the expression or activity of GCREC.

Therefore, in one embodiment, GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a

cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, 5 congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic 10 carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired 15 immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and 20 thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an 25 autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes 30 mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, 35 systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura,

ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, 5 hepatitis virus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togaviruses.

In another embodiment, a vector capable of expressing GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those described above.

10 In a further embodiment, a composition comprising a substantially purified GCREC in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those provided above.

15 In still another embodiment, an agonist which modulates the activity of GCREC may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those listed above.

20 In a further embodiment, an antagonist of GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, described above. In one aspect, an antibody which specifically binds GCREC may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express GCREC.

25 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC including, but not limited to, those described above.

30 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

35 An antagonist of GCREC may be produced using methods which are generally known in the art. In particular, purified GCREC may be used to produce antibodies or to screen libraries of

pharmaceutical agents to identify those which specifically bind GCREC. Antibodies to GCREC may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GCREC or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GCREC have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of GCREC amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to GCREC may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GCREC-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as

disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for GCREC may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between GCREC and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering GCREC epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GCREC. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of GCREC-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GCREC epitopes, represents the average affinity, or avidity, of the antibodies for GCREC. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GCREC epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10⁹ to 10¹² L/mole are preferred for use in immunoassays in which the GCREC-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10⁶ to 10⁷ L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of GCREC, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GCREC-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity,

and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding GCREC, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding GCREC. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding GCREC. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Cli. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding GCREC may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D.

(1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in GCREC expression or regulation causes disease, the expression of 5 GCREC from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in GCREC are treated by constructing mammalian expression vectors encoding GCREC and introducing these vectors by mechanical means into GCREC-deficient cells. Mechanical transfer technologies for 10 use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

15 Expression vectors that may be effective for the expression of GCREC include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). GCREC may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 20 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the 25 ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding GCREC from a normal individual.

30 Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

35 In another embodiment of the invention, diseases or disorders caused by genetic defects with

respect to GCREC expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GCREC under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding GCREC to target cells which have one or more genetic abnormalities with respect to the expression of GCREC. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing GCREC to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with

ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding GCREC to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for GCREC into the alphavirus genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly,

inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, 10 engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GCREC.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, 15 GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

20 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding GCREC. Such DNA sequences may be incorporated into a wide variety of 25 vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' 30 ends of the molecule, or the use of phosphorothioate or 2'-O-methyl rather than phosphodiesterate linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutoxine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous 35 endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding GCREC. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming 5 oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased GCREC expression or activity, a compound which specifically inhibits expression of the 10 polynucleotide encoding GCREC may be therapeutically useful, and in the treatment of disorders associated with decreased GCREC expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GCREC may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method 15 commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a 20 polynucleotide encoding GCREC is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding GCREC are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence 25 of the polynucleotide encoding GCREC. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific 30 polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a 35 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide

sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) *Nat. Biotechnol.* 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of GCREC, antibodies to GCREC, and mimetics, agonists, antagonists, or inhibitors of GCREC.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising GCREC or fragments thereof. For example, liposome preparations

containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GCREC or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GCREC or fragments thereof, antibodies of GCREC, and agonists, antagonists or inhibitors of GCREC, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD₅₀/ED₅₀ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 µg to 100,000 µg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GCREC may be used for the diagnosis of disorders characterized by expression of GCREC, or in assays to monitor patients being treated with GCREC or agonists, antagonists, or inhibitors of GCREC. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics.

5 Diagnostic assays for GCREC include methods which utilize the antibody and a label to detect GCREC in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in

10 the art and may be used.

A variety of protocols for measuring GCREC, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GCREC expression. Normal or standard values for GCREC expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to

15 GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

20 In another embodiment of the invention, the polynucleotides encoding GCREC may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of GCREC may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess

25 expression of GCREC, and to monitor regulation of GCREC levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GCREC or closely related molecules may be used to identify nucleic acid sequences which encode GCREC. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a

30 conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding GCREC, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the GCREC encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:24-46 or from

genomic sequences including promoters, enhancers, and introns of the GCREC gene.

Means for producing specific hybridization probes for DNAs encoding GCREC include the cloning of polynucleotide sequences encoding GCREC or GCREC derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may 5 be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding GCREC may be used for the diagnosis of disorders 10 associated with expression of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in 15 particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other 20 extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann- 25 Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, 30 peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial 35 frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis,

hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

5 degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis,

10 15 passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS),

20 25 Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis,

30 35 erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and

extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, 5 picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus. The polynucleotide sequences encoding GCREC may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered GCREC expression. Such qualitative or quantitative methods are well known in the art.

10 In a particular aspect, the nucleotide sequences encoding GCREC may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GCREC may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a 15 standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding GCREC in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

20 In order to provide a basis for the diagnosis of a disorder associated with expression of GCREC, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GCREC, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from 25 normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, 30 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or 35 overexpressed) in biopsied tissue from an individual may indicate a predisposition for the

development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

5 Additional diagnostic uses for oligonucleotides designed from the sequences encoding GCREC may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding GCREC, or a fragment of a polynucleotide complementary to the polynucleotide encoding GCREC, and will be employed under optimized conditions for identification of a specific gene or
10 condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic
15 disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause
20 differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the
25 sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

30 Methods which may also be used to quantify the expression of GCREC include radiolabeling or biotinylation of nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of
35 interest is presented in various dilutions and a spectrophotometric or colorimetric response gives

rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used 10 to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, GCREC, fragments of GCREC, or antibodies specific for GCREC 15 may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of 20 gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of 25 transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, 30 biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental 35 compounds. All compounds induce characteristic gene expression patterns, frequently termed

molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical

density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the 5 spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

10 A proteomic profile may also be generated using antibodies specific for GCREC to quantify the levels of GCREC expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed 15 by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor 20 correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such 25 cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological 30 sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological 35 sample containing proteins with the test compound. Proteins from the biological sample are

incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated 5 sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-10 2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding GCREC may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. 15 Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial 20 chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state 25 with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic 30 map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GCREC on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as 35 linkage analysis using established chromosomal markers, may be used for extending genetic maps.

Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely 5 localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

10 In another embodiment of the invention, GCREC, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GCREC and the agent being tested may be measured.

15 Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with GCREC, or fragments thereof, and washed. Bound GCREC is then detected by methods well known in the art. Purified GCREC 20 can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GCREC specifically compete with a test compound for binding 25 GCREC. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GCREC.

30 In additional embodiments, the nucleotide sequences which encode GCREC may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

35 The disclosures of all patents, applications and publications, mentioned above and below,

including U.S. Ser. No. 60/208,834, U.S. Ser. No. 60/206,222, U.S. Ser. No. 60/207,476, U.S. Ser. No. 60/208,861, and U.S. Ser. No. 60/209,868, are expressly incorporated by reference herein.

EXAMPLES

5 I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of 10 phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated 15 using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA 20 libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic 25 oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs 30 were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

35 Plasmids obtained as described in Example I were recovered from host cells by *in vivo*

excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 10 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. 15 Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as 20 the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading 25 frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and 30 programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary 35 structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:24-46. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

30 IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative G-protein coupled receptors were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and 35 S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to

form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been annotated as G-protein coupled receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared

by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

- 5 Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in
- 10 Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences
- 15 were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of GCREC Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:24-46 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:24-46 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified

disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \min\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding GCREC are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous

system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, 5 cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding GCREC. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of GCREC Encoding Polynucleotides

10 Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30
15 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

20 High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 25 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

30 The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the 35 concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by

electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and
5 sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site
10 overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following
15 parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing
20 primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

25 **IX. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:24-46 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06
30 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based
35 hybridization analysis of human genomic DNA digested with one of the following endonucleases:

Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature 5 under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing 10 photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link 15 elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

20 Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. 25 After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is 30 described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X 35 first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40

μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

10 **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, 20 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different 30 fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC 35 computer. The digitized data are displayed as an image where the signal intensity is mapped using a

linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

5 A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

10 Sequences complementary to the GCREC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCREC. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GCREC.

15 To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the GCREC-encoding transcript.

XII. Expression of GCREC

20 Expression and purification of GCREC is achieved using bacterial or virus-based expression systems. For expression of GCREC in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).
25 Antibiotic resistant bacteria express GCREC upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GCREC in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding GCREC by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases.
30 Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 35 et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther.

7:1937-1945.)

In most expression systems, GCREC is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GCREC at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified GCREC obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

GCREC function is assessed by expressing the sequences encoding GCREC at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 µg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GCREC on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GCREC and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected 5 cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding GCREC and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of GCREC Specific Antibodies

GCREC substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to 10 immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the GCREC amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is 15 synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A 20 peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the 25 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-GCREC activity by, for example, binding the peptide or GCREC to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring GCREC Using Specific Antibodies

Naturally occurring or recombinant GCREC is substantially purified by immunoaffinity chromatography using antibodies specific for GCREC. An immunoaffinity column is constructed by covalently coupling anti-GCREC antibody to an activated chromatographic resin, such as 30 CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing GCREC are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GCREC (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt 35 antibody/GCREC binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such

as urea or thiocyanate ion), and GCREC is collected.

XVI. Identification of Molecules Which Interact with GCREC

Molecules which interact with GCREC may include agonists and antagonists, as well as molecules involved in signal transduction, such as G proteins. GCREC, or a fragment thereof, is labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) A fragment of GCREC includes, for example, a fragment comprising one or more of the three extracellular loops, the extracellular N-terminal region, or the third intracellular loop. Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GCREC, washed, and any wells with labeled GCREC complex are assayed. Data obtained using different concentrations of GCREC are used to calculate values for the number, affinity, and association of GCREC with the candidate ligand molecules.

Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech). GCREC may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential GCREC agonists or antagonists may be tested for activation or inhibition of GCREC receptor activity using the assays described in sections XVII and XVIII. Candidate molecules may be selected from known GPCR agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Methods for detecting interactions of GCREC with intracellular signal transduction molecules such as G proteins are based on the premise that internal segments or cytoplasmic domains from an orphan G protein-coupled seven transmembrane receptor may be exchanged with the analogous domains of a known G protein-coupled seven transmembrane receptor and used to identify the G-proteins and downstream signaling pathways activated by the orphan receptor domains (Kobilka, B.K. et al. (1988) Science 240:1310-1316). In an analogous fashion, domains of the orphan receptor may be cloned as a portion of a fusion protein and used in binding assays to demonstrate interactions with specific G proteins. Studies have shown that the third intracellular loop of G protein-coupled seven transmembrane receptors is important for G protein interaction and signal transduction (Conklin, B.R. et al. (1993) Cell 73:631-641). For example, the DNA fragment corresponding to the third intracellular loop of GCREC may be amplified by the polymerase chain reaction (PCR) and subcloned into a fusion vector such as pGEX (Pharmacia Biotech). The construct is transformed into an appropriate bacterial host, induced, and the fusion protein is purified from the

cell lysate by glutathione-Sepharose 4B (Pharmacia Biotech) affinity chromatography.

For in vitro binding assays, cell extracts containing G proteins are prepared by extraction with 50 mM Tris, pH 7.8, 1 mM EGTA, 5 mM MgCl₂, 20 mM CHAPS, 20% glycerol, 10 µg of both aprotinin and leupeptin, and 20 µl of 50 mM phenylmethylsulfonyl fluoride. The lysate is incubated 5 on ice for 45 min with constant stirring, centrifuged at 23,000 g for 15 min at 4°C, and the supernatant is collected. 750 µg of cell extract is incubated with glutathione S-transferase (GST) fusion protein beads for 2 h at 4°C. The GST beads are washed five times with phosphate-buffered saline. Bound G subunits are detected by [³²P]ADP-ribosylation with pertussis or cholera toxins. The reactions are terminated by the addition of SDS sample buffer (4.6% (w/v) SDS, 10% (v/v) 10 β-mercaptoethanol, 20% (w/v) glycerol, 95.2 mM Tris-HCl, pH 6.8, 0.01% (w/v) bromphenol blue). The [³²P]ADP-labeled proteins are separated on 10% SDS-PAGE gels, and autoradiographed. The separated proteins in these gels are transferred to nitrocellulose paper, blocked with blotto (5% nonfat 15 dried milk, 50 mM Tris-HCl (pH 8.0), 2 mM CaCl₂, 80 mM NaCl, 0.02% NaN₃, and 0.2% Nonidet P-40) for 1 hour at room temperature, followed by incubation for 1.5 hours with Gα subtype selective antibodies (1:500; Calbiochem-Novabiochem). After three washes, blots are incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit immunoglobulin (1:2000, Cappel, Westchester PA) and visualized by the chemiluminescence-based ECL method (Amersham Corp.).

XVII. Demonstration of GCREC Activity

An assay for GCREC activity measures the expression of GCREC on the cell surface. cDNA 20 encoding GCREC is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using GCREC-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) 25 and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of GCREC expressed on the cell surface.

In the alternative, an assay for GCREC activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding GCREC is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The 30 transiently transfected cells are then incubated in the presence of [³H]thymidine, a radioactive DNA precursor molecule. Varying amounts of GCREC ligand are then added to the cultured cells. Incorporation of [³H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the 35 amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GCREC ligand concentration range is indicative of receptor activity. One unit of activity per

milliliter is defined as the concentration of GCREC producing a 50% response level, where 100% represents maximal incorporation of [³H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

5 In a further alternative, the assay for GCREC activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) J. Biol. Chem. 273:4990-4996). A plasmid encoding full length GCREC is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are 10 grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand 15 compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing 1×10^5 cells/well and incubated with inositol-free media and [³H]myoinositol, 2 μ Ci/well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl 20 followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

25 **XVIII. Identification of GCREC Ligands**

GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence 30 of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca^{2+} . These may be measured directly using standard methods well known in the art, or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green 35 fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) Trends Pharmacol. Sci. 17:235-237). Assay technologies are available for both of these second messenger systems to allow high

throughput readout in multi-well plate format, such as the adenylyl cyclase activation FlashPlate Assay (NEN Life Sciences Products), or fluorescent Ca²⁺ indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, GCREC may be coexpressed with the G-proteins G_{α15/16} which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel the signal transduction of the GCREC through a pathway involving phospholipase C and Ca²⁺ mobilization. Alternatively, GCREC may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCREC activation screening. These yeast systems substitute a human GPCR and G_α protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways are also modified so that the normal yeast response to the signal is converted to positive growth on selective media or to reporter gene expression (Broach, J.R. and J. Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incye Project ID	Polypeptide SEQ ID NO:	Incye Polypeptide ID	Polynucleotide SEQ ID NO:	Incye Polynucleotide ID
7475208	1	7475208CD1	24	7475208CB1
7475101	2	7475101CD1	25	7475101CB1
7475152	3	7475152CD1	26	7475152CB1
7475164	4	7475164CD1	27	7475164CB1
7475170	5	7475170CD1	28	7475170CB1
7475197	6	7475197CD1	29	7475197CB1
7475210	7	7475210CD1	30	7475210CB1
7475221	8	7475221CD1	31	7475221CB1
7475244	9	7475244CD1	32	7475244CB1
7475293	10	7475293CD1	33	7475293CB1
7475297	11	7475297CD1	34	7475297CB1
7475193	12	7475193CD1	35	7475193CB1
7475213	13	7475213CD1	36	7475213CB1
7475272	14	7475272CD1	37	7475272CB1
7475200	15	7475200CD1	38	7475200CB1
7475121	16	7475121CD1	39	7475121CB1
7475165	17	7475165CD1	40	7475165CB1
7475273	18	7475273CD1	41	7475273CB1
7476077	19	7476077CD1	42	7476077CB1
7476113	20	7476113CD1	43	7476113CB1
7476117	21	7476117CD1	44	7476117CB1
7476079	22	7476079CD1	45	7476079CB1
7476112	23	7476112CD1	46	7476112CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	7475208CD1	g12745520	7.00E-91	Putative sweet taste receptor TIR1 [Mus musculus]
		g683747	4.00E-73	Extracellular calcium-sensing receptor [Homo sapiens]
2	7475101CD1	g1256389	5.70E-95	Taste bud receptor protein TB 334 [Rattus norvegicus] (Thomas, M.B. et al. (1996) Gene 178:1-5)
3	7475152CD1	g2370145	2.30E-82	Olfactory receptor protein [Homo sapiens] (Bernot, A. et al. (1997) Nat. Genet. 17:25-31)
4	7475164CD1	g11692559	1.00E-141	Odorant receptor K42 [Mus musculus]
5	7475170CD1	g12054409	1.00E-107	Olfactory receptor [Homo sapiens]
6	7475197CD1	g2808658	1.60E-90	Olfactory receptor [Homo sapiens] (Bernot, A. et al. (1998) Genomics 50:147-160)
7	7475210CD1	g1256389	3.90E-135	Taste bud receptor protein TB 334 [Rattus norvegicus] (Thomas, M.B. et al. (1996) Gene 178:1-5)
8	7475221CD1	g6178008	2.10E-104	Odorant receptor MOR18 [Mus musculus] (Tsuboi, A. et al. (1999) J. Neurosci. 19:8409-8418)
9	7475244CD1	g38331598	2.90E-84	Olfactory receptor [Homo sapiens] (Buettnner, J.A. et al. (1998) Genomics 53:56-68)
10	7475293CD1	g6090787	2.10E-104	Olfactory receptor [Pan troglodytes] (Sharon, D. et al. (1999) Genomics 61:24-36)
11	7475297CD1	g6178008	3.60E-100	Odorant receptor MOR18 [Mus musculus] (Tsuboi, A. et al. (1999) J. Neurosci. 19:8409-8418)
12	7475193CD1	g6178006	4.60E-84	Odorant receptor MOR83 [Mus musculus] (Tsuboi, A. et al. (1999) J. Neurosci. 19:8409-8418)
13	7475213CD1	g1419016	9.70E-139	Odorant receptor [Mus musculus] (Asai, H. et al. (1996) Biochem. Biophys. Res. Commun. 221:240-247)
14	7475272CD1	g3746448	4.70E-75	Olfactory receptor OR93Gib [Hylobates lar] (Rouquier, S. et al. (1998) Hum. Mol. Genet. 7:1337-1345)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
15	7475200CD1	g6178008	4.60E-138	Odorant receptor MOR18 [Mus musculus] (Tsuboi, A. et al. (1999) J. Neurosci. 19:8409-8418)
16	7475121CD1	g3983392	2.70E-85	Olfactory receptor F6 [Mus musculus] (Krautwurst, D. et al. (1998) Cell 95:917-926)
17	7475165CD1	g7211257	1.20E-109	Olfactory receptor [Gorilla gorilla] (Rouquier, S. et al. (2000) Proc. Natl. Acad. Sci. USA 97:2870-2874)
18	7475273CD1	g1314663	4.10E-82	CFOLF2 [Canis familiaris] (Issel-Tarver, L. and J. Rine (1996) Proc. Natl. Acad. Sci. USA 93:10879-10902)
19	7476077CD1	g6532001	1.40E-88	Odorant receptor S19 [Mus musculus]
20	7476113CD1	g1336041	9.30E-92	HsOLF1 [Homo sapiens]
21	7476117CD1	g1336041	2.50E-82	HsOLF1 [Homo sapiens]
22	7476079CD1	g12704541	1.00E-126	Olfactory receptor S83 [Mus musculus]
23	7476112CD1	g3983392	4.00E-100	Olfactory receptor F6 [Mus musculus] (Krautwurst, D. et al. (1998) Cell 95:917-926)

Table 3

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7475208CD1	855	S203 S217 S242 S308 S312 S477 S539 S562 S570 S678 S744 T102 T153 T480 T852	N130 N283 N304 N411 N432 N475 N85	G-PROTEIN COUPLED RECEPTORS FAMILY 3 DM00837 159362 1-893 : N411-E841 G-protein coupled receptor BL00979I: P506-H526	BLAST-DOMO BLIMPS-BLOCKS
2	7475101CD1	330	T25 S84 T285 S308 S324	N22 N82	Metabotropic glutamate receptor signature PR00248: K32-G44, G69-N84, N84-C103, V141-P167, L202-Q221, Q221-V237, V237-F254, A692-P715	BLIMPS-PRINTS

Transmembrane domain:
L581-F601, L617-F635, A692-L711

G-protein coupled receptors family
3 signature 2: C528-C552

Transmembrane domains:
P42-L64; I109-M135; L214-F233

7 transmembrane receptor (rhodopsin family) domain: G58-Y307

G-protein coupled receptors
signature BL00237: Q107-P146; L224-Y235; I299-K315

G-protein coupled receptors
signature: Y119-V164

Olfactory receptor signature
PR00245: M76-K97; F194-D208;
F255-G270; A291-L302; S308-F322

Rhodopsin-like GPCR superfamily
signature PR00237: L43-S67;
M76-K97; L121-I143; L157-L178;
I216-F239; A254-L278; S289-K315

RECEPTOR OLFACTORY RECEPTORLIKE
GPROTEIN COUPLED TRANSMEMBRANE
GLYCOPROTEIN MULTIGENE FAMILY
PD000921: V183-L262

G-PROTEIN COUPLED RECEPTORS
DM000131 P23266 | 17-306: L34-L321

G-protein coupled receptors motif:
L127-I143

MOTIFS

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7475152CD1	324	S19 S67 S93 T267 S18 S87 S290 S315 T318	N5 N276	Signal peptide: M1-S21 Transmembrane domain: L30-I46 7 transmembrane receptor (rhodopsin family) domain: G41-Y289 G-protein coupled receptors signature BL00237: K90-P129; V207-Y218; T281-K297 G-protein coupled receptors signature: Y102-F147 Olfactory receptor signature PR00245: M59-K80; F177-S191; F238-G253; A273-L284; S290-I304 Rhodopsin-like GPCR superfamily signature PR00237: P26-H50; M59-K80; F104-I126; A199-L222; R271-K297 RECEPTOR OLFACTORY RECEPTOR LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY FD000921: L166-I245 G-PROTEIN COUPLED RECEPTORS DM00013 P23266 17-306: L17-I304 G-protein coupled receptors motif: I110-I126	HMMER HMMER HMMER - PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-DOMO MOTIFS HMMER HMMER - PFAM BLIMPS-BLOCKS PROFILESCAN
4	7475164CD1	374	T368 T44 S130 S156 T179 T329 S14 T81 T141 S200 T223 S354		Transmembrane domains: F91-L111; I260-I279 7 transmembrane receptor (rhodopsin family) domain: G104-I265; S338-Y353 G-protein coupled receptors signature BL00237: N153-P192; I345-K361 G-protein coupled receptors signature: Y165-S213	

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4					Olfactory receptor signature PR00245: V122-K143; Y240-S254; F301-G316; S337-L348; S354-T368 Rhodopsin-like GPCR superfamily signature PR00237: P89-A113; V122-K143; F167-I189; L262-F285; K335-K361	BLIMPS-PRINTS
5	7475170CD1	312	S49 S67 T193 S18 T291	N5 N42 N65 N195 N265	G-PROTEIN COUPLED RECEPTORS DM00013 S51356 18-307: Transmembrane domains: L23-G41; M59-L82; C97-M118; F200-F216 7 transmembrane receptor (rhodopsin family) domain: G41-Y290 G-protein coupled receptors signature BL00237: K90-P129; L207-Y218; T282-K298 Olfactory receptor signature PR00245: M59-Q80; F177-D191; F238-G253; I274-I285; T291-L305 OLFACTOORY RECEPTOR RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: T246-Y309 G-PROTEIN COUPLED RECEPTORS DM00013 P23275 17-306: S18-L305	BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	7475197CD1	325	S323 T21 S80 S201 T278 T283 S304	N18 N78 N144	Signal peptide: M1-G54 Transmembrane domains: L43-I59; V211-F229 7 transmembrane receptor (rhodopsin family) domain: G54-Y303 G-protein coupled receptors signature BL00237: Q103-P142; I220-Y231; T295-K311 Olfactory receptor signature PR00245: M72-K93; F190-D204; F251-G266; G287-I298; S304-I318	SPSCAN HMMER
7	7475210CD1	311	S6 S65 S186 S289 S304	N3 N63	RECEPTOR OLFACTORY RECEPTOR LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L179-L258 G-PROTEIN COUPLED RECEPTORS DM00013 P23266 17-306: K32-I318 Transmembrane domains: I28-I44; M195-T214 7 transmembrane receptor (rhodopsin family) domain: G39-Y288 G-protein coupled receptors signature BL00237: H88-P127; L205-Y216; T280-K296 G-protein coupled receptors signature: Y100-L145 Olfactory receptor signature PR00245: M57-K78; F175-D189; F236-G251; A272-L283; S289-F303 Rhodopsin-like GPCR superfamily signature PR00237: M57-K78; F102-I124; V138-F159; V197-V220; A235-C259; I270-K296	HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L164-L243 G-PROTEIN COUPLED RECEPTORS DM00013 P23266 17-306: I15-S304 G-protein coupled receptors motif: L108-I124	BLAST-PRODOM
8	7475221CD1	344	S335 T25 S95 S115 S252 T316 S331	N36 N290	Transmembrane domain: V54-V75 7 transmembrane receptor (rhodopsin family) domain: G69-Y315 G-protein coupled receptors BL00237: K118-P157; E259-L285; T307-K323 G-protein coupled receptors F130-A175 Olfactory receptor signature PR00245: M87-K108; F205-N219; F265-V280; M299-L310; T316-W330 Rhodopsin-like GPCR superfamily signature PR00237: V54-M78; M87-K108; D132-I154; V168-L189; M227-L250; A264-R288; K297-K323 RECEPTOR OLFACTORY RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L194-V272 G-PROTEIN COUPLED RECEPTORS DM00013 S29710 15-301: L45-W330 G-protein coupled receptors motif: A138-I154	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	7475244CD1	313	S68 S168 S189 S3 T79 S138 S196 S233 S292	N6	Transmembrane domains: F29-T149; I93-M119; L199-T225 7 transmembrane receptor (rhodopsin family) domain: G42-Y291 G-protein coupled receptors signature BL00237: R91-P130; I283-N299 G-protein coupled receptors signature: F104-G153 Olfactory receptor signature PR00245: M60-K81; F178-D192; F239-G254; A275-L286; S292-V306 RECEPTOR OLFACTORY RECEPTOR-LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L167-L246 G-PROTEIN COUPLED RECEPTORS DM00013 S51316 18-307: S19-V307 G-protein coupled receptors motif: T111-V127 Transmembrane domains: L30-I46; V198-I216 7 transmembrane receptor (rhodopsin family) domain: G41-Y290 G-protein coupled receptors signature BL00237: Q90-P129; I207-Y218; T282-K298 G-protein coupled receptors signature: Y102-V147 Olfactory receptor signature PR00245: M59-K80; F177-D191; L238-G253; A274-L285; S291-F305 RECEPTOR OLFACTORY RECEPTOR-LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-L245	HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS HMMER HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM
10	7475293CD1	313	S8 T108 S188 S193 S268 S230 S268 S291	N5	Transmembrane domains: L30-I46; V198-I216 7 transmembrane receptor (rhodopsin family) domain: G41-Y290 G-protein coupled receptors signature BL00237: Q90-P129; I207-Y218; T282-K298 G-protein coupled receptors signature: Y102-V147 Olfactory receptor signature PR00245: M59-K80; F177-D191; L238-G253; A274-L285; S291-F305 RECEPTOR OLFACTORY RECEPTOR-LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-L245	HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM

Table 3 (cont.)

SEQ NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites		Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10						G-PROTEIN COUPLED RECEPTORS DM00013 P30953 18-306: P18-N306	BLAST-DOMO
11	7475297CD1	309	T36 S65 S52 S91 S135 S222 S227 T286	N6		G-protein coupled receptors motif: L110-I126 Signal peptide: M1-R54 Transmembrane domains: V28-V44; M57-A76; M204-L220 7 transmembrane receptor (rhodopsin family) domain: E39-Y285	MOTIFS SPSCAN HMMER
						G-protein coupled receptors signature BL00237: T88-P127; T277-K293	BLIMPS-BLOCKS
						G-protein coupled receptors signature: F100-G144	PROFILESCAN
						Olfactory receptor signature PR00245: M57-K78; F175-D189; L235-V250; M269-L280; T286-W300	BLIMPS-PRINTS
						Rhodopsin-like GPCR superfamily signature PR00237: I24-I48; M57-K78; E102-I124; V138-L159; V197-L220; A234-R258; K267-K293	BLIMPS-PRINTS
						RECEPTOR OLFACTORY RECEPTOR-LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: I164-L242	BLAST-PRODOM
						G-PROTEIN COUPLED RECEPTORS DM00013 S29710 15-301: L15-W300	BLAST-DOMO
						G-protein coupled receptors motif: V108-I124	MOTIFS
12	7475193CD1	313	S229 T77 T192 S148 T235 T290	N5		Transmembrane domains: V26-I45; I200-A219	HMMER
						7 transmembrane receptor (rhodopsin family) domain: G41-Y289	HMMER-PFAM
						G-protein coupled receptors signature BL00237: K90-P129; F281-K297	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12					Olfactory receptor signature PR00245: M59-E80; Y177-N191; M239-G254; V273-R284; T290-V304 Rhodopsin-like GPCR superfamily signature PR00237: V26-S50; M59-E80; I104-I126; K271-K297 OLFACTORY RECEPTOR RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD194621: T247-V304 G-PROTEIN COUPLED RECEPTORS DM00013 S29710 15-301: L17-L303 Transmembrane domains: L27-C50; I196-L219 7 transmembrane receptor (rhodopsin family) domain: A41-Y289	BLIMPS-PRINTS
13	7475213CD1	342	T236 T171 S187 T192 S265 S309 S290	N5	G-protein coupled receptors signature BL00237: Q90-P129; I206-Y217; T281-Q297 G-protein coupled receptors signature: F102-G147 Olfactory receptor signature PR00245: M59-R80; F176-D190; F237-G252; L273-L284; S290-L304 RECEPTOR OLFACTORY RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-L244 G-PROTEIN COUPLED RECEPTORS DM00013 P30954 29-316: S18-L300	BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7475272CD1	310	S172 T188 S267 S290	N5	Signal peptide: M1-G41 Transmembrane domains: F28-L48; F202-M226 7 transmembrane receptor (rhodopsin family) domain: G41-Y289 G-protein coupled receptors signature BL00237: A90-P129; I281-K297 G-protein coupled receptors signature: F102-A146 Olfactory receptor signature PR00245: F237-G252; V273-L284; S290-L304 Rhodopsin-like GPCR superfamily signature PR00237: P26-L50; M59-Q80; F104-V126; I199-I222; R271-K297	HMMER-PFAM SPSCAN HMMER
15	7475200CD1	302	S222 S65 S83 T286 Y85	N130 N6 N63	OLFACOTRY RECEPTOR RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: T245-R306 G-PROTEIN COUPLED RECEPTORS DM00013 S51356 18-307: G-protein coupled receptors motif: I110-V126 signal cleavage: M1-A54 transmembrane domain: V27-L53, L196-L223 7 transmembrane receptor (rhodopsin family) 7tm_1: G39-Y285 G-protein coupled receptor BL00237A: R88-P127, BL00237D: T277-K293	BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS SPSCAN HMMER HMMER-PFAM BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15					Olfactory receptor signature PR00245A: V57-K78, PR00245B: F175-N189, PR00245C: L235-V250, PR00245D: M269-L280, PR00245E: T286-F300 Rhodopsin-like GPCR superfamily signature PR00237A: V24-T48, PR00237B: V57-K78, PR00237C: A102-I124, PR00237D: L138-L159, PR00237E: V197-L220, PR00237F: A234-H258, PR00237G: K267-K293 G-protein coupled receptors signature: A102-V145	BLIMPS-PRINTS
16	7475121CD1	316	S68, T79, S138, S293	N5, N192	G-Protein Receptor: V108-I124 G-PROTEIN COUPLED RECEPTORS DM00013 S29710 15-301: L15-F300, DM00013 P23266 17-306: L15-L299, DM00013 P37067 17-306: L15-L299, DM00013 P23270 18-311: V24-K298 RECEPTOR OLFACTORY RECEPTOR LIKE G- PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L164-I243 G-PROTEIN COUPLED RECEPTORS: DM00013 P30954 29-316: S18-I303 OLFACTORY RECEPTOR-LIKE G-PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN, MULTIGENE FAMILY: PD000921: L167-L247 G-protein coupled receptor: BL00237A: H91-P130; BL00237C: T284-K300	MOTIFS BLAST-DOMO BLAST-DOMO BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16					Olfactory receptor signature: PR00245A: M60-R81; PR00245B: F178-N192; PR00245C: F240-S255; PR00245D: M276-L287; PR00245E: S293-F307	BLIMPS-PRINTS
17	7475165CD1	370	S125 S288 S349 S364 S57 T225 T228 T35 T46 T52 Y152	N123 N63	EDG1 orphan receptor signature: PR00642D: T49-F63 G-protein coupled receptors signature: F103-T149 Transmembrane domain: I27-L45, M102-Y121, V204-V224 7-Transmembrane receptor (rhodopsin family: 7tm_1): G42-F292	BLIMPS-PRINTS PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites		Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7475273CD1	318	S65, T84, S135, S186, S266, S289, S298, T316	N3, N144		G-PROTEIN COUPLED RECEPTORS: DM00013 S51356 18-307; T16-M299 OLFACTORY RECEPTOR-LIKE G-PROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN, MULTIGENE FAMILY: PD149621:T244-K305	BLAST-DOMO BLAST-PRODOM
19	7476077CD1	321	S231 S69 T179 T263 T7	N44 N5		G-protein coupled receptor: BL00237A: K88-P127; BL00237D: I280-K296 Olfactory receptor signature: PR00245A: M57-N78; PR00245B: V175-D189; PR00245C: F236-G251; PR00245D: V272-L283; PR00245E: S289-F303 EDG1 Orphan receptor signature: PRO0642D: V46-F60 Transmembrane (transmem_domain):T23-V46; I90-M116; I195-I221 7-transmembrane receptor motif (rhodopsin family; 'tm_1'): G39-V138; I209-Y288	BLIMPS-BLOCKS BLIMPS-PRINTS HMMER BLIMPS-PRINTS HMMER-PFAM HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Domains and Motifs	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19					Olfactory receptor signature PR00245; V61-K82; T179-D193; L240-T255 RA1C PD170483; V249-A319 G-PROTEIN COUPLED RECEPTORS DM00013 G45774 18-309; G-protein coupled receptors motif: M112-I128	BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS	BLIMPS-PRINTS
20	7476113CD1	313	S138 S189 S233 S292 S68 T205 T271 T301 T4	N136 N37 N7	Transmembrane domains: F29-V48; F102-D122 7 transmembrane receptor (rhodopsin family) domain: G42-Y291 G-protein coupled receptors signature BL00237; R91-P130; I283-K299	HMMER HMMER-PPFAM BLIMPS-BLOCKS PROFILESCAN signature: F103-V147	HMMER-PPFAM BLIMPS-PRINTS BLIMPS-BLOCKS PROFILESCAN
21	7476117CD1	328	S139 S190 S293 S69 T206 T227 T272 T4 T8	N7	Receptor olfactory receptorlike gprotein coupled transmembrane glycoprotein multigene family PD000921; L167-L246 G-PROTEIN COUPLED RECEPTORS DM00013 S51356 18-307; P22-K299 G-protein coupled receptors motif: T111-I127	BLAST-PRODOM BLAST-DOMO MOTIFS	BLIMPS-PRINTS HMMER HMMER-PPFAM BLIMPS-BLOCKS

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21					Olfactory receptor signature PR00245: M61-M82; F179-D193; F240-G255; A276-L287; S293-T307 RECEPTOR OLFACTORY RECEPTOR-LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L168-L247	BLIMPS-PRINTS
22	7476079CD1	324	S102 S13 S179 S7	N12	G-PROTEIN COUPLED RECEPTORS DM00013 S51356 18-307: E24-T1303 Signal peptide: M1-A49 Transmembrane domains: L40-I57; L75-W95; P142-V165; L211-T230; H253-T273 7 transmembrane receptor (rhodopsin family) domain: A50-T146 G-protein coupled receptors signature BL00237: K99-P138; P292-R308 G-protein coupled receptors signature: Y111-L159 Olfactory receptor signature PR00245: M68-K89; C186-D200; L247-T262 Melanocortin receptor family signature PR00534: Q60-L72; I135-T146; T304-A317 G-PROTEIN COUPLED RECEPTORS DM00013 G45774 18-309: G-protein coupled receptors motif: M119-I135	BLAST-PRODOM SPSCAN HMMER

Table 3 (cont.)

SEQ ID NO: 23	Incyte Polypeptide ID: 7476112CD1	Amino Acid Residues 315	Potential Phosphorylation Sites S137 S292 S51 S67 S8 T142 T88	Potential Glycosylation Sites N5	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Transmembrane domains: M26-L44; L61-I78; A150-F168; L202-I229	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G41-F291	HMMER-PFAM
					G-protein coupled receptors signature BL00237: R90-P129; T283-K299	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F102-C147	PROFILES CAN
					Olfactory receptor signature PR00245: M59-R80; F177-D191; F239-G254; M275-I286; S292-C306	BLIMPS-PRINTS
					Melanocortin receptor family signature PR00534: S51-L63; I126-S137; V200-F212	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTOR-LIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-L246	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013 P30954 29-316: S18-M302	BLAST-DOMO

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
24	7475208CB1	2739	1276-1513, 1-1200,	7669623H1 (NOSEDIC02) GNN.g7523967_000013_002	2123 1	2739 2602
			1622-2183, 2299-2335, 2463-2739			
25	7475101CB1	993	252-993, 1-149	GNN.g7329615_000006_002	1	993
26	7475152CB1	990	1-27, 919-990, 777-819	GNN.g7329615_000004_002	1	990
27	7475164CB1	1125	470-1011, 1084-1125, 58-396	GNN.g3738097_004	1	1125
28	7475170CB1	939	1-30, 20-939	GNN.g6453999_000016_004	1	939
29	7475197CB1	978	1-872, 921-978	GNN.g7024166_000032_004	1	978
30	7475210CB1	936	1-112, 195-936	GNN.g7329615_000007_002	1	936
31	7475221CB1	1035	1-89, 1002-1035, 760-910	GNN.g7321527_000008_004	1	1035
32	7475244CB1	942	1-339, 396-942	GNN.g6806865_000020_002	1	942
33	7475293CB1	942	1-98, 190- 826, 904- 942	GNN.g7329615_000013_002	1	942
34	7475297CB1	930	1-354, 390-930	GNN.g6806865_000016_002	1	930
35	7475193CB1	942	1-230, 479-942	GNN.g7321521_000022_002 GBI:g7321521_000022.rawcomp	233 1	942 360

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
36	7475213CB1	1029	1-297, 591-1029	GNN.g7134787_000015_002	1	1029
37	7475272CB1	933	1-835, 893-933	GNN.g7024166_000035_002	1	933
38	7475200CB1	948	1-381, 415-948	GNN.g7143464_000027_004	1	948
39	7475121CB1	951	386-951, 1-349	GNN.g6910525_000003_004	1	951
40	7475165CB1	1113	1-210, 418-717, 1068-1113	GNN.g4092817_004 g2525800	1	1113
41	7475273CB1	957	279-298, 416-635, 876-957	GNN.g6984471_000006_002	1	934
42	7476077CB1	966	1-333, 409-966	GNN.g7658497_000015_002	1	957
43	7476113CB1	975	439-852, 1-378, 930-975	GNN.g7705148_000007_002	1	966
44	7476117CB1	987	1-354, 885-987, 411-822	GNN.g7705148_000018_004	1	975
45	7476079CB1	975	1-190, 426-975	GNN.g7658497_000018_002	1	987
46	7476112CB1	948	574-948	GNN.g7690171_000001_002	1	948

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
24	7475208CB1	NOSEDIC02

Table 6

Library	Vector	Library Description
NOSEDIC02	ESP0RT1	This large size fractionated library was constructed using RNA isolated from nasal polyp tissue.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	EST ₃ : Probability value=1.0E-8 or less <i>Full Length sequences</i> : Probability value=1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and search.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	EST ₃ : fasta E value=1.0E-6 Assembled EST ₃ : fasta Identity=95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value=1.0E-3 or less <i>Signal peptide hits</i> : Score=0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater.
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Int'l. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

Tissues	Polynucleotide SEQ ID NO:					
	25	27	28	30	32	33
Breast, Fat, Skin	+	+	+	-	+	+
Muscle, Bone, Synovium, Connective tissue	+	+	+	-	-	+
Pancreas, Liver, Gallbladder	+	+	+	-	+	-
Brain: Amygdala, Thalamus, Hippocampus, Entorhinal cortex, Archaeocortex	+	+	+	+	-	-
Brain: Striatum, Caudate nucleus, Putamen, Dentate nucleus, Globus pallidus, Substantia innominata, Raphe magnus	+	-	-	-	-	-
Kidney, Fetal colon, Small intestine, Ileum, Esophagus	+	+	-	-	-	-
Fetal heart, Aorta, Coronary artery	-	-	-	-	+	-
Fetal lung, Adult lung	+	-	+	-	+	-
Placenta, Prostate, Uterus	-	-	+	-	+	-
Olfactory bulb	+	-	+	-	+	-

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23,
 - b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and
 - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.
 2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-23.
 3. An isolated polynucleotide encoding a polypeptide of claim 1.
 4. An isolated polynucleotide encoding a polypeptide of claim 2.
 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:24-46.
 6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
 7. A cell transformed with a recombinant polynucleotide of claim 6.
 8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
 9. A method for producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
 - a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46,
 - b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46,
 - c) a polynucleotide complementary to a polynucleotide of a),
 - d) a polynucleotide complementary to a polynucleotide of b), and
 - e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
 - a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
 - b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
 - a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
 - b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

18. A method for treating a disease or condition associated with decreased expression of 5 functional GCREC, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- 10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
 b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

15 21. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 20.

20 22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
 b) detecting antagonist activity in the sample.

25 23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

30 24. A method for treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- 5 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 10 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 15 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- 20 c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- 25 a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- 30 c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of GCREC in a biological sample comprising the steps of:

- 5 a) combining the biological sample with an antibody of claim 10, under conditions suitable
for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
b) detecting the complex, wherein the presence of the complex correlates with the presence
of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

- 10 a) a chimeric antibody,
b) a single chain antibody,
c) a Fab fragment,
d) a F(ab')₂ fragment, or
e) a humanized antibody.

15

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of GCREC
in a subject, comprising administering to said subject an effective amount of the composition of claim
20 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of GCREC
25 in a subject, comprising administering to said subject an effective amount of the composition of claim
33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim
10 comprising:

- 30 a) immunizing an animal with a polypeptide having an amino acid sequence selected from
the group consisting of SEQ ID NO:1-23, or an immunogenic fragment thereof, under conditions to
elicit an antibody response;
b) isolating antibodies from said animal; and
c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal

antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

36. An antibody produced by a method of claim 35.

5

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

10 a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-

15 producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

20

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

30

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a

polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the
5 group consisting of SEQ ID NO:1-23 from a sample, the method comprising:
a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

10

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

15 47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

20 50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

25 52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

30 55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
- 5 59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
- 10 62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
- 15 64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.
- 20 67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.
68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.
- 25 69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.
70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.
- 30 71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
5 NO:29.

74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.

10 75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.

15 76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.

20 78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.

79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:35.

25 80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.

81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
30 NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.

83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:39.

84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
5 NO:40.

85. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:41.

10

86. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:42.

15 NO:43.

88. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:44.

20

89. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:45.

25 NO:46.

<110> INCYTE GENOMICS, INC.
PATTERSON, Chandra
TRIBOULEY, Catherine M.
YAO, Monique G.
GRIFFIN, Jennifer A.
THORNTON, Michael
LU, Yan
KALLICK, Deborah A.
GANDHI, Ameena R.
AU-YOUNG, Jamice

<120> G-PROTEIN COUPLED RECEPTORS

<130> SF-0781 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/206,222; 60/207,476; 60/208,834; 60/208,861; 60/209,868
<151> 2000-05-22; 2000-05-25; 2000-06-02; 2000-06-02; 2000-06-07

<160> 46

<170> PERL Program

<210> 1

<211> 855

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475208CD1

<400> 1

Met	Leu	Gly	Pro	Ala	Val	Leu	Gly	Leu	Ser	Leu	Trp	Ala	Leu	Leu
1		5						10					15	
His	Pro	Gly	Thr	Gly	Ala	Pro	Leu	Cys	Leu	Ser	Gln	Gln	Leu	Arg
	20								25				30	
Met	Lys	Gly	Asp	Tyr	Val	Leu	Gly	Gly	Leu	Phe	Pro	Leu	Gly	Glu
	35								40				45	
Ala	Glu	Glu	Ala	Gly	Leu	Arg	Ser	Arg	Thr	Arg	Pro	Ser	Ser	Pro
	50							55					60	
Val	Cys	Thr	Arg	Phe	Ser	Ser	Asn	Gly	Leu	Leu	Trp	Ala	Leu	Ala
	65								70				75	
Met	Lys	Met	Ala	Val	Glu	Glu	Ile	Asn	Asn	Lys	Ser	Asp	Leu	Leu
	80								85				90	
Pro	Gly	Leu	Arg	Leu	Gly	Tyr	Asp	Leu	Phe	Asp	Thr	Cys	Ser	Glu
	95								100				105	
Pro	Val	Val	Ala	Met	Lys	Pro	Ser	Leu	Met	Phe	Leu	Ala	Lys	Ala
	110								115				120	
Gly	Ser	Arg	Asp	Ile	Ala	Ala	Tyr	Cys	Asn	Tyr	Thr	Gln	Tyr	Gln
	125								130				135	
Pro	Arg	Val	Leu	Ala	Val	Ile	Gly	Pro	His	Ser	Ser	Glu	Leu	Ala
	140								145				150	
Met	Val	Thr	Gly	Lys	Phe	Phe	Ser	Phe	Phe	Leu	Met	Pro	Gln	Val
	155								160				165	
Ala	Pro	Pro	Thr	Ile	Thr	His	Pro	His	Pro	Ala	Leu	Pro	Val	Gly
	170								175				180	
Ala	Pro	Val	Ser	Gly	Asp	Ala	Ser	Trp	Pro	Leu	Gln	Val	Ser	Tyr
	185								190				195	
Gly	Ala	Ser	Met	Glu	Leu	Leu	Ser	Ala	Arg	Glu	Thr	Phe	Pro	Ser
	200								205				210	

Phe Phe Arg Thr Val Pro Ser Asp Arg Val Gln Leu Thr Ala Ala
 215 220 225
 Ala Glu Leu Leu Gln Glu Phe Gly Trp Asn Trp Val Ala Ala Leu
 230 235 240
 Gly Ser Asp Asp Glu Tyr Gly Arg Gln Gly Leu Ser Ile Phe Ser
 245 250 255
 Ala Leu Ala Arg His Ala Ala Ser Ala Ser Arg Thr Arg Ala Trp
 260 265 270
 Cys Arg Cys Pro Val Gln Asp Val Leu His Gln Val Asn Gln Ser
 275 280 285
 Ser Val Gln Val Val Leu Leu Phe Ala Ser Val His Ala Ala His
 290 295 300
 Ala Leu Phe Asn Tyr Ser Ile Ser Ser Arg Leu Ser Pro Lys Val
 305 310 315
 Trp Val Ala Ser Glu Ala Trp Leu Thr Ser Asp Leu Val Met Gly
 320 325 330
 Leu Pro Gly Met Ala Gln Met Gly Thr Val Leu Gly Phe Leu Gln
 335 340 345
 Arg Gly Ala Gln Leu His Glu Phe Pro Gln Tyr Val Lys Thr His
 350 355 360
 Leu Ala Leu Ala Thr Asp Pro Ala Phe Cys Ser Ala Leu Gly Glu
 365 370 375
 Arg Glu Gln Gly Leu Glu Glu Asp Val Val Gly Gln Arg Cys Pro
 380 385 390
 Gln Cys Asp Cys Ile Thr Leu Gln Asn Arg Ala Gln Ala Leu His
 395 400 405
 Asn Thr Leu Gln Cys Asn Ala Ser Gly Cys Pro Ala Gln Asp Pro
 410 415 420
 Val Lys Pro Trp Gln Leu Leu Glu Asn Met Tyr Asn Leu Thr Phe
 425 430 435
 His Val Gly Gly Leu Pro Leu Arg Phe Asp Ser Ser Gly Asn Val
 440 445 450
 Asp Met Glu Tyr Asp Leu Lys Leu Trp Val Trp Gln Gly Ser Val
 455 460 465
 Pro Arg Leu His Asp Val Gly Arg Phe Asn Gly Ser Leu Arg Thr
 470 475 480
 Glu Arg Leu Lys Ile Arg Trp His Thr Ser Asp Asn Gln Pro Ser
 485 490 495
 Arg Ala Arg Pro Gln Ala Cys Ala Gln Lys Pro Val Ser Arg Cys
 500 505 510
 Ser Arg Gln Cys Gln Glu Gly Gln Val Arg Arg Val Lys Gly Phe
 515 520 525
 His Ser Cys Cys Tyr Asp Cys Val Asp Cys Glu Ala Gly Ser Tyr
 530 535 540
 Arg Gln Asn Pro Asp Asp Ile Ala Cys Thr Phe Cys Gly Gln Asp
 545 550 555
 Glu Trp Ser Pro Glu Arg Ser Thr Arg Cys Phe Arg Arg Arg Ser
 560 565 570
 Arg Phe Leu Ala Trp Gly Glu Pro Ala Val Leu Leu Leu Leu
 575 580 585
 Leu Leu Ser Leu Ala Leu Gly Leu Val Leu Ala Leu Gly Leu
 590 595 600
 Phe Val His His Arg Asp Ser Pro Leu Val Gln Ala Ser Gly Gly
 605 610 615
 Pro Leu Ala Cys Phe Gly Leu Val Cys Leu Gly Leu Val Cys Leu
 620 625 630
 Ser Val Leu Leu Phe Pro Gly Gln Pro Ser Pro Ala Arg Cys Leu
 635 640 645
 Ala Gln Gln Pro Leu Ser His Leu Pro Leu Thr Gly Cys Leu Ser
 650 655 660
 Thr Leu Phe Leu Gln Ala Ala Glu Ile Phe Val Glu Ser Glu Leu
 665 670 675
 Pro Leu Ser Trp Ala Asp Arg Leu Ser Gly Cys Leu Arg Gly Pro

680	685	690
Trp Ala Trp Leu Val Val Leu Leu Ala	Met Leu Val Glu Val	Ala
695	700	705
Leu Cys Thr Trp Tyr Leu Val Ala Phe Pro Pro Glu Val Val	Thr	
710	715	720
Gly Leu Ala His Ala Ala His Gly Gly Ala Gly Ala Leu Pro His		
725	730	735
Thr Leu Leu Gly Gln Leu Arg Pro Ser Ala Arg His His Ala Thr		
740	745	750
Leu Ala Phe Leu Cys Phe Thr Gly His Phe Pro Gly Ala Glu Pro		
755	760	765
Ala Gly Pro Leu Gln Pro Cys His Val Ala Ser His Ile Cys His		
770	775	780
Ala Gly Leu Leu His His Thr Gly Ser His Phe Val Pro Leu Leu		
785	790	795
Ala Gln Cys Ala Gly Gly His Ser Gly Pro Ala Val Gln Met Gly		
800	805	810
Ala Leu Leu Leu Cys Val Leu Gly Ile Leu Ala Ala Phe His Leu		
815	820	825
Pro Arg Cys Tyr Leu Leu Met Arg Gln Pro Gly Leu Asn Thr Pro		
830	835	840
Glu Phe Phe Leu Gly Gly Pro Gly Asp Ala Thr Arg Pro Glu		
845	850	855

<210> 2
<211> 330
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475101CD1

<400> 2		
Met Glu Gly Phe Tyr Leu Arg Arg Ser His Glu Leu Gln Gly Met		
1	5	10
Gly Lys Pro Gly Arg Val Asn Gln Thr Thr Val Ser Asp Phe Leu		
20	25	30
Leu Leu Gly Leu Ser Glu Trp Pro Glu Glu Gln Pro Leu Leu Phe		
35	40	45
Gly Ile Phe Leu Gly Met Tyr Leu Val Thr Met Val Gly Asn Leu		
50	55	60
Leu Ile Ile Leu Ala Ile Ser Ser Asp Pro His Leu His Thr Pro		
65	70	75
Met Tyr Phe Phe Leu Ala Asn Leu Ser Leu Thr Asp Ala Cys Phe		
80	85	90
Thr Ser Ala Ser Ile Pro Lys Met Leu Ala Asn Ile His Thr Gln		
95	100	105
Ser Gln Ile Ile Ser Tyr Ser Gly Cys Leu Ala Gln Leu Tyr Phe		
110	115	120
Leu Leu Met Phe Gly Gly Leu Asp Asn Cys Leu Leu Ala Val Met		
125	130	135
Ala Tyr Asp Arg Tyr Val Ala Ile Cys Gln Pro Leu His Tyr Ser		
140	145	150
Thr Ser Met Ser Pro Gln Leu Cys Ala Leu Met Leu Gly Val Cys		
155	160	165
Trp Val Leu Thr Asn Cys Pro Ala Leu Met His Thr Leu Leu Leu		
170	175	180
Thr Arg Val Ala Phe Cys Ala Gln Lys Ala Ile Pro His Phe Tyr		
185	190	195
Cys Asp Pro Ser Ala Leu Leu Lys Leu Ala Cys Ser Asp Thr His		
200	205	210

Val	Asn	Glu	Leu	Met	Ile	Ile	Thr	Met	Gly	Leu	Leu	Phe	Leu	Thr	
215									220					225	
Val	Pro	Leu	Leu	Leu	Ile	Ile	Val	Phe	Ser	Tyr	Val	Arg	Ile	Phe	Trp
230									235						240
Ala	Val	Phe	Val	Ile	Ser	Ser	Pro	Gly	Gly	Arg	Trp	Lys	Ala	Phe	
245									250						255
Ser	Thr	Cys	Gly	Ser	His	Leu	Thr	Val	Val	Leu	Leu	Phe	Tyr	Gly	
260									265						270
Ser	Leu	Met	Gly	Val	Tyr	Leu	Leu	Pro	Pro	Ser	Thr	Tyr	Ser	Thr	
275									280						285
Glu	Arg	Glu	Ser	Arg	Ala	Ala	Val	Leu	Tyr	Met	Val	Ile	Ile	Pro	
290									295						300
Thr	Leu	Asn	Pro	Phe	Ile	Tyr	Ser	Leu	Arg	Asn	Arg	Asp	Met	Lys	
305									310						315
Glu	Ala	Leu	Gly	Lys	Leu	Phe	Val	Ser	Gly	Lys	Thr	Phe	Phe	Leu	
320									325						330

<210> 3
<211> 324
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475152CD1

<400>	3														
Met	Gly	Met	Ser	Asn	Leu	Thr	Arg	Leu	Ser	Glu	Phe	Ile	Leu	Leu	
1									10					15	
Gly	Leu	Ser	Ser	Arg	Ser	Glu	Asp	Gln	Arg	Pro	Leu	Phe	Ala	Leu	
									20					30	
Phe	Leu	Ile	Ile	Tyr	Leu	Val	Thr	Leu	Met	Gly	Asn	Leu	Leu	Ile	
									35					45	
Ile	Leu	Ala	Ile	His	Ser	Asp	Pro	Arg	Leu	Gln	Asn	Pro	Met	Tyr	
									50					60	
Phe	Phe	Leu	Ser	Ile	Leu	Ser	Phe	Ala	Asp	Ile	Cys	Tyr	Thr	Thr	
									65					75	
Val	Ile	Val	Pro	Lys	Met	Leu	Val	Asn	Phe	Leu	Ser	Glu	Lys	Lys	
									80					90	
Thr	Ile	Ser	Tyr	Ala	Glu	Cys	Leu	Ala	Gln	Met	Tyr	Phe	Phe	Leu	
									95					105	
Val	Phe	Gly	Asn	Ile	Asp	Ser	Tyr	Leu	Leu	Ala	Ala	Met	Ala	Ile	
									110					120	
Asn	Arg	Cys	Val	Ala	Ile	Cys	Asn	Pro	Phe	His	Tyr	Val	Thr	Val	
									125					135	
Met	Asn	Arg	Arg	Cys	Cys	Val	Leu	Leu	Ala	Phe	Pro	Ile	Thr		
									140					150	
Phe	Ser	Tyr	Phe	His	Ser	Leu	Leu	His	Val	Leu	Leu	Val	Asn	Arg	
									155					165	
Leu	Thr	Phe	Cys	Thr	Ser	Asn	Val	Ile	His	His	Phe	Phe	Cys	Asp	
									170					180	
Val	Asn	Pro	Val	Leu	Lys	Leu	Ser	Cys	Ser	Ser	Thr	Phe	Val	Asn	
									185					195	
Glu	Ile	Val	Ala	Met	Thr	Glu	Gly	Leu	Ala	Ser	Val	Met	Ala	Pro	
									200					210	
Phe	Val	Cys	Ile	Ile	Ile	Ser	Tyr	Leu	Arg	Ile	Leu	Ile	Ala	Val	
									215					225	
Leu	Lys	Ile	Pro	Ser	Ala	Ala	Gly	Lys	His	Lys	Ala	Phe	Ser	Thr	
									230					240	
Cys	Ser	Ser	His	Leu	Thr	Val	Val	Ile	Leu	Phe	Tyr	Gly	Ser	Ile	
									245					255	
Ser	Tyr	Val	Tyr	Leu	Gln	Pro	Leu	Ser	Ser	Tyr	Thr	Val	Lys	Asp	

260	265	270
Arg Ile Ala Thr Ile Asn Tyr Thr Val	Leu Thr Ser Val Leu	Asn
275	280	285
Pro Phe Ile Tyr Ser Leu Arg Asn Lys	Asp Met Lys Arg Gly	Leu
290	295	300
Gln Lys Leu Ile Asn Lys Ile Lys Ser	Gln Met Ser Arg Phe	Ser
305	310	315
Thr Lys Thr Asn Lys Ile Cys Gly Pro		
320		

<210> 4

<211> 374

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475164CD1

<400> 4

Met Ala Ile Cys Asn Pro Leu Leu Tyr Asn Ile Ala Met Ser Pro		
1 5 10 15		
Lys Val Cys Ser Ser His Met Leu Gly Ser Tyr Phe Trp Pro Phe		
20 25 30		
Ser Gly Ala Met Ala His Thr Arg Cys Met Leu Lys Leu Thr Ser		
35 40 45		
Cys Glu Ala Asn Thr Ile Asn His Tyr Phe Cys Asp Thr Leu His		
50 55 60		
Leu Leu Gln Leu Ser Cys Thr Ser Thr Tyr Val Arg Ala Glu Phe		
65 70 75		
Ile Leu Ala Gly Leu Thr Gln Arg Pro Glu Leu Gln Leu Pro Leu		
80 85 90		
Phe Leu Leu Phe Leu Gly Ile Tyr Val Val Thr Val Val Gly Asn		
95 100 105		
Leu Gly Met Ile Phe Leu Ile Ala Leu Ser Ser Gln Leu Tyr Pro		
110 115 120		
Pro Val Tyr Tyr Phe Leu Ser His Leu Ser Phe Ile Asp Leu Cys		
125 130 135		
Tyr Ser Ser Val Ile Thr Pro Lys Met Leu Val Asn Phe Val Pro		
140 145 150		
Glu Glu Asn Ile Ile Ser Phe Leu Glu Cys Ile Thr Gln Leu Tyr		
155 160 165		
Phe Phe Leu Ile Phe Val Ile Ala Glu Gly Tyr Leu Leu Thr Ala		
170 175 180		
Met Glu Tyr Asp Arg Tyr Val Ala Ile Cys Arg Pro Leu Leu Tyr		
185 190 195		
Asn Ile Val Met Ser His Arg Val Cys Ser Ile Met Met Ala Val		
200 205 210		
Val Tyr Ser Leu Gly Phe Leu Trp Ala Thr Val His Thr Thr Arg		
215 220 225		
Met Ser Val Leu Ser Phe Cys Arg Ser His Thr Val Ser His Tyr		
230 235 240		
Phe Cys Asp Ile Leu Pro Leu Leu Thr Leu Ser Cys Ser Ser Thr		
245 250 255		
His Ile Asn Glu Ile Leu Leu Phe Ile Ile Gly Gly Val Asn Thr		
260 265 270		
Leu Ala Thr Thr Leu Ala Val Leu Ile Ser Tyr Ala Phe Ile Phe		
275 280 285		
Ser Ser Ile Leu Gly Ile His Ser Thr Glu Gly Gln Ser Lys Ala		
290 295 300		
Phe Gly Thr Cys Ser Ser His Leu Leu Ala Val Gly Ile Phe Phe		
305 310 315		
Gly Ser Ile Thr Phe Met Tyr Phe Lys Pro Pro Ser Ser Thr Thr		

320	325	330
Met Glu Lys Glu Lys Val Ser Ser Val Phe Tyr Ile Thr Ile Ile		
335	340	345
Pro Met Leu Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val		
350	355	360
Lys Asn Ala Leu Lys Lys Met Thr Arg Gly Arg Gln Ser Ser		
365	370	

<210> 5

<211> 312

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475170CD1

<400> 5

Met Asp Gln Lys Asn Gly Ser Ser Phe Thr Gly Phe Ile Leu Leu			
1	5	10	15
Gly Phe Ser Asp Arg Pro Gln Leu Glu Leu Val Leu Phe Val Val			
20	25	30	
Leu Leu Ile Phe Tyr Ile Phe Thr Leu Leu Gly Asn Lys Thr Ile			
35	40	45	
Ile Val Leu Ser His Leu Asp Pro His Leu His Thr Pro Met Tyr			
50	55	60	
Phe Phe Phe Ser Asn Leu Ser Phe Leu Asp Leu Cys Tyr Thr Thr			
65	70	75	
Gly Ile Val Pro Gln Leu Leu Val Asn Leu Arg Gly Ala Asp Lys			
80	85	90	
Ser Ile Ser Tyr Gly Gly Cys Val Val Gln Leu Tyr Ile Ser Leu			
95	100	105	
Gly Leu Gly Ser Thr Glu Cys Val Leu Leu Gly Val Met Val Phe			
110	115	120	
Asp Arg Tyr Ala Ala Val Cys Arg Pro Leu His Tyr Thr Val Val			
125	130	135	
Met His Pro Cys Leu Tyr Val Leu Met Ala Ser Thr Ser Trp Val			
140	145	150	
Ile Gly Phe Ala Asn Ser Leu Leu Gln Thr Val Leu Ile Leu Leu			
155	160	165	
Leu Thr Leu Cys Gly Arg Asn Lys Leu Glu His Phe Leu Cys Glu			
170	175	180	
Val Pro Pro Leu Leu Lys Leu Ala Cys Val Asp Thr Thr Met Asn			
185	190	195	
Glu Ser Glu Leu Phe Phe Val Ser Val Ile Ile Leu Leu Val Pro			
200	205	210	
Val Ala Leu Ile Ile Phe Ser Tyr Ser Gln Ile Val Arg Ala Val			
215	220	225	
Met Arg Ile Lys Leu Ala Thr Gly Gln Arg Lys Val Phe Gly Thr			
230	235	240	
Cys Gly Ser His Leu Thr Val Val Ser Leu Phe Tyr Gly Thr Ala			
245	250	255	
Ile Tyr Ala Tyr Leu Gln Pro Gly Asn Asn Tyr Ser Gln Asp Gln			
260	265	270	
Gly Lys Phe Ile Ser Leu Phe Tyr Thr Ile Ile Thr Pro Met Ile			
275	280	285	
Asn Pro Leu Ile Tyr Thr Leu Arg Asn Lys Asp Val Lys Gly Ala			
290	295	300	
Leu Lys Lys Val Leu Trp Lys Asn Tyr Asp Ser Arg			
305	310		

<210> 6

<211> 325

<212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475197CD1

<400> 6

Met	Lys	Thr	Phe	Ser	Ser	Phe	Leu	Gln	Ile	Gly	Arg	Asn	Met	His
1									10					15
Gln	Gly	Asn	Gln	Thr	Thr	Ile	Thr	Glu	Phe	Ile	Leu	Leu	Gly	Phe
									20	25				30
Phe	Lys	Gln	Asp	Glu	His	Gln	Asn	Leu	Leu	Phe	Val	Leu	Phe	Leu
									35	40				45
Gly	Met	Tyr	Leu	Val	Thr	Val	Ile	Gly	Asn	Gly	Leu	Ile	Ile	Val
									50	55				60
Ala	Ile	Ser	Leu	Asp	Thr	Tyr	Leu	His	Thr	Pro	Met	Tyr	Leu	Phe
									65	70				75
Leu	Ala	Asn	Leu	Ser	Phe	Ala	Asp	Ile	Ser	Ser	Ile	Ser	Asn	Ser
									80	85				90
Val	Pro	Lys	Met	Leu	Val	Asn	Ile	Gln	Thr	Lys	Ser	Gln	Ser	Ile
									95	100				105
Ser	Tyr	Glu	Ser	Cys	Ile	Thr	Gln	Met	Tyr	Phe	Ser	Ile	Val	Phe
									110	115				120
Val	Val	Ile	Asp	Asn	Leu	Leu	Leu	Gly	Thr	Met	Ala	Tyr	Asp	His
									125	130				135
Phe	Val	Ala	Ile	Cys	His	Pro	Leu	Asn	Tyr	Thr	Ile	Leu	Met	Arg
									140	145				150
Pro	Arg	Phe	Gly	Ile	Leu	Leu	Thr	Val	Ile	Ser	Trp	Phe	Leu	Ser
									155	160				165
Asn	Ile	Ile	Ala	Leu	Thr	His	Thr	Leu	Leu	Leu	Ile	Gln	Leu	Leu
									170	175				180
Phe	Cys	Asn	His	Asn	Thr	Leu	Pro	His	Phe	Phe	Cys	Asp	Leu	Ala
									185	190				195
Pro	Leu	Leu	Lys	Leu	Ser	Cys	Ser	Asp	Thr	Leu	Ile	Asn	Glu	Leu
									200	205				210
Val	Leu	Phe	Ile	Val	Gly	Leu	Ser	Val	Ile	Ile	Phe	Pro	Phe	Thr
									215	220				225
Leu	Ser	Phe	Phe	Ser	Tyr	Val	Cys	Ile	Ile	Arg	Ala	Val	Leu	Arg
									230	235				240
Val	Ser	Ser	Thr	Gln	Gly	Lys	Trp	Lys	Ala	Phe	Ser	Thr	Cys	Gly
									245	250				255
Ser	His	Leu	Thr	Val	Val	Leu	Leu	Phe	Tyr	Gly	Thr	Ile	Val	Gly
									260	265				270
Val	Tyr	Phe	Phe	Pro	Ser	Ser	Thr	His	Pro	Glu	Asp	Thr	Asp	Lys
									275	280				285
Ile	Gly	Ala	Val	Leu	Phe	Thr	Val	Val	Thr	Pro	Met	Ile	Asn	Pro
									290	295				300
Phe	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Asp	Met	Lys	Gly	Ala	Leu	Arg
									305	310				315
Lys	Leu	Ile	Asn	Arg	Lys	Ile	Ser	Ser	Leu					
									320	325				

<210> 7
 <211> 311
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475210CD1

<400> 7

Met	Glu	Asn	Gln	Ser	Ser	Ile	Ser	Glu	Phe	Phe	Leu	Arg	Gly	Ile
1								10						15
Ser	Ala	Pro	Pro	Glu	Gln	Gln	Gln	Ser	Leu	Phe	Gly	Ile	Phe	Leu
									25					30
Cys	Met	Tyr	Leu	Val	Thr	Leu	Thr	Gly	Asn	Leu	Leu	Ile	Ile	Leu
									40					45
Ala	Ile	Gly	Ser	Asp	Leu	His	Leu	His	Thr	Pro	Met	Tyr	Phe	Phe
									55					60
Leu	Ala	Asn	Leu	Ser	Phe	Val	Asp	Met	Gly	Leu	Thr	Ser	Ser	Thr
									70					75
Val	Thr	Lys	Met	Leu	Val	Asn	Ile	Gln	Thr	Arg	His	His	Thr	Ile
									85					90
Ser	Tyr	Thr	Gly	Cys	Leu	Thr	Gln	Met	Tyr	Phe	Phe	Leu	Met	Phe
									100					105
Gly	Asp	Leu	Asp	Ser	Phe	Phe	Leu	Ala	Ala	Met	Ala	Tyr	Asp	Arg
									115					120
Tyr	Val	Ala	Ile	Cys	His	Pro	Leu	Cys	Tyr	Ser	Thr	Val	Met	Arg
									130					135
Pro	Gln	Val	Cys	Ala	Leu	Met	Leu	Ala	Leu	Cys	Trp	Val	Leu	Thr
									145					150
Asn	Ile	Val	Ala	Leu	Thr	His	Thr	Phe	Leu	Met	Ala	Arg	Leu	Ser
									160					165
Phe	Cys	Val	Thr	Gly	Glu	Ile	Ala	His	Phe	Phe	Cys	Asp	Ile	Thr
									175					180
Pro	Val	Leu	Lys	Leu	Ser	Cys	Ser	Asp	Thr	His	Ile	Asn	Glu	Met
									190					195
Met	Val	Phe	Val	Leu	Gly	Gly	Thr	Val	Leu	Ile	Val	Pro	Phe	Leu
									205					210
Cys	Ile	Val	Thr	Ser	Tyr	Ile	His	Ile	Val	Pro	Ala	Ile	Leu	Arg
									220					225
Val	Arg	Thr	Arg	Gly	Gly	Val	Gly	Lys	Ala	Phe	Ser	Thr	Cys	Ser
									235					240
Ser	His	Leu	Cys	Val	Val	Cys	Val	Phe	Tyr	Gly	Thr	Leu	Phe	Ser
									250					255
Ala	Tyr	Leu	Cys	Pro	Pro	Ser	Ile	Ala	Ser	Glu	Glu	Lys	Asp	Ile
									265					270
Ala	Ala	Ala	Ala	Met	Tyr	Thr	Ile	Val	Thr	Pro	Met	Leu	Asn	Pro
									275					285
Phe	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Asp	Met	Lys	Gly	Ala	Leu	Lys
									290					300
Arg	Leu	Phe	Ser	His	Arg	Ser	Ile	Val	Ser	Ser				
									305					310

<210> 8

<211> 344

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475221CD1

<400> 8

Met	Glu	Leu	Leu	Thr	Asn	Asn	Leu	Lys	Phe	Ile	Thr	Asp	Pro	Phe
1									10					15
Val	Cys	Arg	Leu	Arg	His	Leu	Ser	Pro	Thr	Pro	Ser	Glu	Glu	His
									25					30
Met	Lys	Asn	Lys	Asn	Asn	Val	Thr	Glu	Phe	Ile	Leu	Leu	Gly	Leu
									35					45
Thr	Gln	Asn	Pro	Glu	Gly	Gln	Lys	Val	Leu	Phe	Val	Thr	Phe	Leu
									50					60
Leu	Ile	Tyr	Met	Val	Thr	Ile	Met	Gly	Asn	Leu	Leu	Ile	Ile	Val
									65					75

Thr Ile Met Ala Ser Gln Ser Leu Gly Ser Pro Met Tyr Phe Phe
 80 85 90
 Leu Ala Ser Leu Ser Phe Ile Asp Thr Val Tyr Ser Thr Ala Phe
 95 100 105
 Ala Pro Lys Met Ile Val Asp Leu Leu Ser Glu Lys Lys Thr Ile
 110 115 120
 Ser Phe Gln Gly Cys Met Ala Gln Leu Phe Met Asp His Leu Phe
 125 130 135
 Ala Gly Ala Glu Val Ile Leu Leu Val Val Met Ala Tyr Asp Arg
 140 145 150
 Tyr Met Ala Ile Cys Lys Pro Leu His Glu Leu Ile Thr Met Asn
 155 160 165
 Arg Arg Val Cys Val Leu Met Leu Leu Ala Ala Trp Ile Gly Gly
 170 175 180
 Phe Leu His Ser Leu Val Gln Phe Leu Phe Ile Tyr Gln Leu Pro
 185 190 195
 Phe Cys Gly Pro Asn Val Ile Asp Asn Phe Leu Cys Asp Leu Tyr
 200 205 210
 Pro Leu Leu Lys Leu Ala Cys Thr Asn Thr Tyr Val Thr Gly Leu
 215 220 225
 Ser Met Ile Ala Asn Gly Gly Ala Ile Cys Ala Val Thr Phe Phe
 230 235 240
 Thr Ile Leu Leu Ser Tyr Gly Val Ile Leu His Ser Leu Lys Thr
 245 250 255
 Gln Ser Leu Glu Gly Lys Arg Lys Ala Phe Tyr Thr Cys Ala Ser
 260 265 270
 His Val Thr Val Val Ile Leu Phe Phe Val Pro Cys Ile Phe Leu
 275 280 285
 Tyr Ala Arg Pro Asn Ser Thr Phe Pro Ile Asp Lys Ser Met Thr
 290 295 300
 Val Val Leu Thr Phe Ile Thr Pro Met Leu Asn Pro Leu Ile Tyr
 305 310 315
 Thr Leu Lys Asn Ala Glu Met Lys Ser Ala Met Arg Lys Leu Trp
 320 325 330
 Ser Lys Lys Val Ser Leu Ala Gly Lys Trp Leu Tyr His Ser
 335 340

<210> 9
 <211> 313
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475244CD1

<400> 9

Met Ala Ser Glu Arg Asn Gln Ser Ser Thr Pro Thr Phe Ile Leu		
1 5 10 15		
Leu Gly Phe Ser Glu Tyr Pro Glu Ile Gln Val Pro Leu Phe Leu		
20 25 30		
Val Phe Leu Phe Val Tyr Thr Val Thr Val Val Gly Asn Leu Gly		
35 40 45		
Met Ile Ile Ile Ile Arg Leu Asn Ser Lys Leu His Thr Ile Met		
50 55 60		
Tyr Phe Phe Leu Ser His Leu Ser Leu Thr Asp Phe Cys Phe Ser		
65 70 75		
Thr Val Val Thr Pro Lys Leu Leu Glu Asn Leu Val Val Glu Tyr		
80 85 90		
Arg Thr Ile Ser Phe Ser Gly Cys Ile Met Gln Phe Cys Phe Ala		
95 100 105		
Cys Ile Phe Gly Val Thr Glu Thr Phe Met Leu Ala Ala Met Ala		
110 115 120		

Tyr	Asp	Arg	Phe	Val	Ala	Val	Cys	Lys	Pro	Leu	Leu	Tyr	Thr	Thr
				125				130				135		
Ile	Met	Ser	Gln	Lys	Leu	Cys	Ala	Leu	Leu	Val	Ala	Gly	Ser	Tyr
				140				145				150		
Thr	Trp	Gly	Ile	Val	Cys	Ser	Leu	Ile	Leu	Thr	Tyr	Phe	Leu	Leu
				155				160				165		
Asp	Leu	Ser	Phe	Cys	Glu	Ser	Thr	Phe	Ile	Asn	Asn	Phe	Ile	Cys
				170				175				180		
Asp	His	Ser	Val	Ile	Val	Ser	Ala	Ser	Tyr	Ser	Asp	Pro	Tyr	Ile
				185				190				195		
Ser	Gln	Arg	Leu	Cys	Phe	Ile	Ile	Ala	Ile	Phe	Asn	Glu	Val	Ser
				200				205				210		
Ser	Leu	Ile	Ile	Ile	Leu	Thr	Ser	Tyr	Met	Leu	Ile	Phe	Thr	Thr
				215				220				225		
Ile	Met	Lys	Met	Arg	Ser	Ala	Ser	Gly	Arg	Gln	Lys	Thr	Phe	Ser
				230				235				240		
Thr	Cys	Ala	Ser	His	Leu	Thr	Ala	Ile	Thr	Ile	Phe	His	Gly	Thr
				245				250				255		
Ile	Leu	Phe	Leu	Tyr	Cys	Val	Pro	Asn	Pro	Lys	Thr	Ser	Ser	Leu
				260				265				270		
Ile	Val	Thr	Val	Ala	Ser	Val	Phe	Tyr	Thr	Val	Ala	Ile	Pro	Met
				275				280				285		
Leu	Asn	Pro	Leu	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Asp	Ile	Asn	Asn
				290				295				300		
Met	Phe	Glu	Lys	Leu	Val	Val	Thr	Lys	Leu	Ile	Tyr	His		
				305				310						

<210> 10

<211> 313

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475293CD1

<400> 10

Met	Lys	Arg	Glu	Asn	Gln	Ser	Ser	Val	Ser	Glu	Phe	Leu	Leu	Leu
1				5				10				15		
Asp	Leu	Pro	Ile	Trp	Pro	Glu	Gln	Gln	Ala	Val	Phe	Phe	Thr	Leu
				20				25				30		
Phe	Leu	Gly	Met	Tyr	Leu	Ile	Thr	Val	Leu	Gly	Asn	Leu	Ile	
				35				40				45		
Ile	Leu	Leu	Ile	Arg	Leu	Asp	Ser	His	Leu	His	Thr	Pro	Met	Phe
				50				55				60		
Phe	Phe	Leu	Ser	His	Leu	Ala	Leu	Thr	Asp	Ile	Ser	Leu	Ser	Ser
				65				70				75		
Val	Thr	Val	Pro	Lys	Met	Leu	Leu	Ser	Met	Gln	Thr	Gln	Asp	Gln
				80				85				90		
Ser	Ile	Leu	Tyr	Ala	Gly	Cys	Val	Thr	Gln	Met	Tyr	Phe	Phe	Ile
				95				100				105		
Phe	Phe	Thr	Asp	Leu	Asp	Asn	Phe	Leu	Leu	Thr	Ser	Met	Ala	Tyr
				110				115				120		
Asp	Arg	Tyr	Val	Ala	Ile	Cys	His	Pro	Leu	Arg	Tyr	Thr	Thr	Ile
				125				130				135		
Met	Lys	Glu	Gly	Leu	Cys	Asn	Leu	Leu	Val	Thr	Val	Ser	Trp	Ile
				140				145				150		
Leu	Ser	Cys	Thr	Asn	Ala	Leu	Ser	His	Thr	Leu	Leu	Leu	Ala	Gln
				155				160				165		
Leu	Ser	Phe	Cys	Ala	Asp	Asn	Thr	Ile	Pro	His	Phe	Phe	Cys	Asp
				170				175				180		
Leu	Val	Ala	Leu	Leu	Lys	Leu	Ser	Cys	Ser	Asp	Ile	Ser	Leu	Asn
				185				190				195		

Glu	Leu	Val	Ile	Phe	Thr	Val	Gly	Gln	Ala	Val	Ile	Thr	Leu	Pro
200									205					210
Leu	Ile	Cys	Ile	Leu	Ile	Ser	Tyr	Gly	His	Ile	Gly	Val	Thr	Ile
215									220					225
Leu	Lys	Ala	Pro	Ser	Thr	Lys	Gly	Ile	Phe	Lys	Ala	Leu	Ser	Thr
230									235					240
Cys	Gly	Ser	His	Leu	Ser	Val	Val	Ser	Leu	Tyr	Tyr	Gly	Thr	Ile
245									250					255
Ile	Gly	Leu	Tyr	Phe	Leu	Pro	Ser	Ser	Ser	Ala	Ser	Ser	Asp	Lys
260									265					270
Asp	Val	Ile	Ala	Ser	Val	Met	Tyr	Thr	Val	Ile	Thr	Pro	Leu	Leu
275									280					285
Asn	Pro	Phe	Ile	Tyr	Ser	Leu	Arg	Asn	Arg	Asp	Ile	Lys	Gly	Ala
290									295					300
Leu	Glu	Arg	Leu	Phe	Asn	Arg	Ala	Thr	Val	Leu	Ser	Gln		
									310					

<210> 11
<211> 309
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475297CD1

Glu	Asn	Gln	Asn	Asn	Val	Thr	Glu	Phe	Ile	Leu	Leu	Gly	Leu	
1									10				15	
Met														
Thr	Glu	Asn	Leu	Glu	Leu	Trp	Lys	Ile	Phe	Ser	Ala	Val	Phe	
20									25					30
Val	Met	Tyr	Val	Ala	Thr	Val	Leu	Glu	Asn	Leu	Leu	Ile	Val	
35									40					45
Thr	Ile	Ile	Thr	Ser	Gln	Ser	Leu	Arg	Ser	Pro	Met	Tyr	Phe	
50									55					60
Leu	Thr	Phe	Leu	Ser	Leu	Leu	Asp	Val	Met	Phe	Ser	Ser	Val	
65									70					75
Ala	Pro	Lys	Val	Ile	Val	Asp	Thr	Leu	Ser	Lys	Ser	Thr	Thr	
80									85					90
Ser	Leu	Lys	Gly	Cys	Leu	Thr	Gln	Leu	Phe	Val	Glu	His	Phe	
95									100					105
Gly	Gly	Val	Gly	Ile	Ile	Leu	Leu	Thr	Val	Met	Ala	Tyr	Asp	
110									115					120
Tyr	Val	Ala	Ile	Cys	Lys	Pro	Leu	His	Tyr	Thr	Ile	Ile	Ser	
125									130					135
Pro	Arg	Val	Cys	Cys	Leu	Met	Val	Gly	Gly	Ala	Trp	Val	Gly	
140									145					150
Phe	Met	His	Ala	Met	Ile	Gln	Leu	Leu	Phe	Met	Tyr	Gln	Ile	
155									160					165
Phe	Cys	Gly	Pro	Asn	Ile	Ile	Asp	His	Phe	Ile	Cys	Asp	Leu	
170									175					180
Gln	Leu	Leu	Thr	Leu	Ala	Cys	Thr	Asp	Thr	His	Ile	Leu	Gly	
185									190					195
Leu	Val	Thr	Leu	Asn	Ser	Gly	Met	Met	Cys	Val	Ala	Ile	Phe	
200									205					210
Ile	Leu	Ile	Ala	Ser	Tyr	Thr	Val	Ile	Leu	Cys	Ser	Leu	Lys	
215									220					225
Tyr	Ser	Ser	Lys	Gly	Arg	His	Lys	Ala	Leu	Ser	Thr	Cys	Ser	
230									235					240
His	Leu	Thr	Val	Val	Val	Leu	Phe	Phe	Val	Pro	Cys	Ile	Phe	
245									250					255
Tyr	Met	Arg	Pro	Val	Val	Thr	His	Pro	Ile	Asp	Lys	Ala	Met	
260									265					270

Val Ser Asp Ser Ile Ile Thr Pro Met Leu Asn Pro Leu Ile Tyr
 275 280 285
 Thr Leu Arg Asn Ala Glu Val Lys Ser Ala Met Lys Lys Leu Trp
 290 295 300
 Met Lys Trp Glu Ala Leu Ala Gly Lys
 305

<210> 12
<211> 313
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475193CD1

<400> 12

Met	Glu	Thr	Ala	Asn	Tyr	Thr	Lys	Val	Thr	Glu	Phe	Val	Leu	Thr
1					5				10					15
Gly	Leu	Ser	Gln	Thr	Pro	Glu	Val	Gln	Leu	Val	Leu	Phe	Val	Ile
									20	25				30
Phe	Leu	Ser	Phe	Tyr	Leu	Phe	Ile	Leu	Pro	Gly	Asn	Ile	Leu	Ile
									35	40				45
Ile	Cys	Thr	Ile	Ser	Leu	Asp	Pro	His	Leu	Thr	Ser	Pro	Met	Tyr
									50	55				60
Phe	Leu	Leu	Ala	Asn	Leu	Ala	Phe	Leu	Asp	Ile	Trp	Tyr	Ser	Ser
									65	70				75
Ile	Thr	Ala	Pro	Glu	Met	Leu	Ile	Asp	Phe	Phe	Val	Glu	Arg	Lys
									80	85				90
Ile	Ile	Ser	Phe	Asp	Gly	Cys	Ile	Ala	Gln	Leu	Phe	Phe	Leu	His
									95	100				105
Phe	Ala	Gly	Ala	Ser	Glu	Met	Phe	Leu	Leu	Thr	Val	Met	Ala	Phe
									110	115				120
Asp	Leu	Tyr	Thr	Ala	Ile	Cys	Arg	Pro	Leu	His	Tyr	Ala	Thr	Ile
									125	130				135
Met	Asn	Gln	Arg	Leu	Cys	Cys	Ile	Leu	Val	Ala	Leu	Ser	Trp	Arg
									140	145				150
Gly	Gly	Phe	Ile	His	Ser	Ile	Ile	Gln	Val	Ala	Leu	Ile	Val	Arg
									155	160				165
Leu	Pro	Phe	Cys	Gly	Pro	Asn	Glu	Leu	Asp	Ser	Tyr	Phe	Cys	Asp
									170	175				180
Ile	Thr	Gln	Val	Val	Arg	Ile	Ala	Cys	Ala	Asn	Thr	Phe	Pro	Glu
									185	190				195
Glu	Leu	Val	Met	Ile	Cys	Ser	Ser	Gly	Leu	Ile	Ser	Val	Val	Cys
									200	205				210
Leu	Ile	Ala	Leu	Leu	Met	Ser	Tyr	Ala	Phe	Leu	Leu	Ala	Leu	Phe
									215	220				225
Lys	Lys	Leu	Ser	Gly	Ser	Gly	Glu	Asn	Thr	Asn	Arg	Ala	Met	Ser
									230	235				240
Thr	Cys	Tyr	Ser	His	Ile	Thr	Ile	Val	Val	Leu	Met	Phe	Gly	Pro
									245	250				255
Ser	Ile	Tyr	Ile	Tyr	Ala	Arg	Pro	Phe	Asp	Ser	Phe	Ser	Leu	Asp
									260	265				270
Lys	Val	Val	Ser	Val	Phe	Asn	Thr	Leu	Ile	Phe	Pro	Leu	Arg	Asn
									275	280				285
Pro	Ile	Ile	Tyr	Thr	Leu	Arg	Asn	Lys	Glu	Val	Lys	Ala	Ala	Met
									290	295				300
Arg	Lys	Leu	Val	Thr	Lys	Tyr	Ile	Leu	Cys	Lys	Glu	Lys		
									305	310				

<210> 13
<211> 342
<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475213CD1

<400> 13

Met	Lys	Arg	Lys	Asn	Phe	Thr	Glu	Val	Ser	Glu	Phe	Ile	Phe	Leu
1									10					15
Gly	Phe	Ser	Ser	Phe	Gly	Lys	His	Gln	Ile	Thr	Leu	Phe	Val	Val
										25				30
Phe	Leu	Thr	Val	Tyr	Ile	Leu	Thr	Leu	Val	Ala	Asn	Ile	Ile	
														45
									40					
Val	Thr	Ile	Ile	Cys	Ile	Asp	His	His	Leu	His	Thr	Pro	Met	Tyr
										55				60
Phe	Phe	Leu	Ser	Met	Leu	Ala	Ser	Ser	Glu	Thr	Val	Tyr	Thr	Leu
										70				75
Val	Ile	Val	Pro	Arg	Met	Leu	Leu	Ser	Leu	Ile	Phe	His	Asn	Gln
										85				90
Pro	Ile	Ser	Leu	Ala	Gly	Cys	Ala	Thr	Gln	Met	Phe	Phe	Phe	Val
										100				105
Ile	Leu	Ala	Thr	Asn	Asn	Cys	Phe	Leu	Leu	Thr	Ala	Met	Gly	Tyr
										115				120
Asp	Arg	Tyr	Val	Ala	Ile	Cys	Arg	Pro	Leu	Arg	Tyr	Thr	Val	Ile
										130				135
Met	Ser	Lys	Gly	Leu	Cys	Ala	Gln	Leu	Val	Cys	Gly	Ser	Phe	Gly
										145				150
Ile	Gly	Leu	Thr	Met	Ala	Val	Leu	His	Val	Thr	Ala	Met	Phe	Asn
										160				165
Leu	Pro	Phe	Cys	Gly	Thr	Val	Val	Asp	His	Phe	Phe	Cys	Asp	Ile
										175				180
Tyr	Pro	Val	Met	Lys	Leu	Ser	Cys	Ile	Asp	Thr	Thr	Ile	Asn	Glu
										190				195
Ile	Ile	Asn	Tyr	Gly	Val	Ser	Ser	Phe	Val	Ile	Phe	Val	Pro	Ile
										205				210
Gly	Leu	Ile	Phe	Ile	Ser	Tyr	Val	Leu	Val	Ile	Ser	Ser	Ile	Leu
										220				225
Gln	Ile	Ala	Ser	Ala	Glu	Gly	Arg	Lys	Lys	Thr	Phe	Ala	Thr	Cys
										235				240
Val	Ser	His	Leu	Thr	Val	Val	Ile	Val	His	Cys	Gly	Cys	Ala	Ser
										250				255
Ile	Ala	Tyr	Leu	Lys	Pro	Lys	Ser	Ser	Ile	Glu	Lys	Asp		
										265				270
Leu	Val	Leu	Ser	Val	Thr	Tyr	Thr	Ile	Ile	Thr	Pro	Leu	Leu	Asn
										280				285
Pro	Val	Val	Tyr	Ser	Leu	Arg	Asn	Lys	Glu	Ile	Gln	Glu	Ser	Leu
										295				300
Gln	Ala	Gly	Leu	Arg	Leu	Leu	Val	Ser	Val	Leu	Glu	Asp	Phe	Ser
										310				315
Phe	Glu	Ser	Phe	Leu	Ala	Pro	Ile	Leu	Pro	Glu	Leu	Ser	Asp	Ser
										325				330
Gln	Ile	Phe	Glu	Leu	Val	Trp	Leu	Gly	Asp	Val	Glu			
										340				

<210> 14

<211> 310

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475272CD1

<400> 14

Met	Ala	Glu	Met	Asn	Leu	Thr	Leu	Val	Thr	Glu	Phe	Leu	Leu	Ile
1									10					15
Ala	Phe	Thr	Glu	Tyr	Pro	Glu	Trp	Ala	Leu	Pro	Leu	Phe	Leu	Leu
				20					25					30
Leu	Leu	Phe	Met	Tyr	Leu	Ile	Thr	Val	Leu	Gly	Asn	Leu	Glu	Met
									35	40				45
Ile	Ile	Ile	Ile	Leu	Met	Asp	His	Gln	Leu	His	Ala	Pro	Met	Tyr
					50				55					60
Phe	Leu	Leu	Ser	His	Leu	Ala	Phe	Met	Asp	Val	Cys	Tyr	Ser	Ser
					65				70					75
Ile	Thr	Val	Pro	Gln	Met	Leu	Ala	Val	Leu	Leu	Glu	His	Gly	Ala
					80				85					90
Ala	Leu	Ser	Tyr	Thr	Arg	Cys	Ala	Ala	Gln	Phe	Phe	Leu	Phe	Thr
					95				100					105
Phe	Phe	Gly	Ser	Ile	Asp	Cys	Tyr	Leu	Leu	Ala	Leu	Met	Ala	Tyr
					110				115					120
Asp	Arg	Tyr	Leu	Ala	Val	Cys	Gln	Pro	Leu	Leu	Tyr	Val	Thr	Ile
					125				130					135
Leu	Thr	Gln	Gln	Ala	Arg	Leu	Ser	Leu	Val	Ala	Gly	Ala	Tyr	Val
					140				145					150
Ala	Gly	Leu	Ile	Ser	Ala	Leu	Val	Arg	Thr	Val	Ser	Ala	Phe	Thr
					155				160					165
Leu	Ser	Phe	Cys	Gly	Thr	Ser	Glu	Ile	Asp	Phe	Ile	Phe	Cys	Asp
					170				175					180
Leu	Pro	Pro	Leu	Leu	Lys	Leu	Thr	Cys	Gly	Glu	Ser	Tyr	Thr	Gln
					185				190					195
Glu	Val	Leu	Ile	Ile	Met	Phe	Ala	Ile	Phe	Val	Ile	Pro	Ala	Ser
					200				205					210
Met	Val	Val	Ile	Leu	Val	Ser	Tyr	Leu	Phe	Ile	Ile	Val	Ala	Ile
					215				220					225
Met	Gly	Ile	Pro	Ala	Gly	Ser	Gln	Ala	Lys	Thr	Phe	Ser	Thr	Cys
					230				235					240
Thr	Ser	His	Leu	Thr	Ala	Val	Ser	Leu	Phe	Phe	Gly	Thr	Leu	Ile
					245				250					255
Phe	Met	Tyr	Leu	Arg	Gly	Asn	Ser	Asp	Gln	Ser	Ser	Glu	Lys	Asn
					260				265					270
Arg	Val	Val	Ser	Val	Leu	Tyr	Thr	Glu	Val	Ile	Pro	Met	Leu	Asn
					275				280					285
Pro	Leu	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Glu	Val	Lys	Glu	Ala	Leu
					290				295					300
Arg	Lys	Ile	Leu	Asn	Arg	Ala	Lys	Leu	Ser					
					305				310					

<210> 15

<211> 302

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475200CD1

<400> 15

Met	Asp	Ile	Pro	Gln	Asn	Ile	Thr	Glu	Phe	Phe	Met	Leu	Gly	Leu
1						5			10					15
Ser	Gln	Asn	Ser	Glu	Val	Gln	Arg	Val	Leu	Phe	Val	Val	Phe	Leu
					20				25					30
Leu	Ile	Tyr	Val	Val	Thr	Val	Cys	Gly	Asn	Met	Leu	Ile	Val	Val
					35				40					45
Thr	Ile	Thr	Ser	Ser	Pro	Thr	Leu	Ala	Ser	Pro	Val	Tyr	Phe	Phe
					50				55					60
Leu	Ala	Asn	Leu	Ser	Phe	Ile	Asp	Thr	Phe	Tyr	Ser	Ser	Met	

65	70	75
Ala Pro Lys Leu Ile Ala Asp Ser Leu Tyr Glu Gly Arg Thr Ile		
80	85	90
Ser Tyr Glu Cys Cys Met Ala Gln Leu Phe Gly Ala His Phe Leu		
95	100	105
Gly Gly Val Glu Ile Ile Leu Leu Thr Val Met Ala Tyr Asp Arg		
110	115	120
Tyr Val Ala Ile Cys Lys Pro Leu His Asn Thr Thr Ile Met Thr		
125	130	135
Arg His Leu Cys Ala Met Leu Val Gly Val Ala Trp Leu Gly Gly		
140	145	150
Phe Leu His Ser Leu Val Gln Leu Leu Leu Val Leu Trp Leu Pro		
155	160	165
Phe Cys Gly Pro Asn Val Ile Asn His Phe Ala Cys Asp Leu Tyr		
170	175	180
Pro Leu Leu Glu Val Ala Cys Thr Asn Thr Tyr Val Ile Gly Leu		
185	190	195
Leu Val Val Ala Asn Ser Gly Leu Ile Cys Leu Leu Asn Phe Leu		
200	205	210
Met Leu Ala Ala Ser Tyr Ile Val Ile Leu Tyr Ser Leu Arg Ser		
215	220	225
His Ser Ala Asp Gly Arg Cys Lys Ala Leu Ser Thr Cys Gly Ala		
230	235	240
His Phe Ile Val Val Ala Leu Phe Phe Val Pro Cys Ile Phe Thr		
245	250	255
Tyr Val His Pro Phe Ser Thr Leu Pro Ile Asp Lys Asn Met Ala		
260	265	270
Leu Phe Tyr Gly Ile Leu Thr Pro Met Leu Asn Pro Leu Ile Tyr		
275	280	285
Thr Leu Arg Asn Glu Glu Val Lys Asn Ala Met Arg Lys Leu Phe		
290	295	300
Thr Trp		

<210> 16
<211> 316
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475121CD1

<400> 16

Met Pro Ser Gln Asn Tyr Ser Ile Ile Ser Glu Phe Asn Leu Phe			
1	5	10	15
Gly Phe Ser Ala Phe Pro Gln His Leu Leu Pro Ile Leu Phe Leu			
20	25	30	
Leu Tyr Leu Leu Met Phe Leu Phe Thr Leu Leu Gly Asn Leu Leu			
35	40	45	
Ile Met Ala Thr Ile Trp Ile Glu His Arg Leu His Thr Pro Met			
50	55	60	
Tyr Leu Phe Leu Cys Thr Leu Ser Val Ser Glu Ile Leu Phe Thr			
65	70	75	
Val Ala Ile Thr Pro Arg Met Leu Ala Asp Leu Leu Ser Thr His			
80	85	90	
His Ser Ile Thr Phe Val Ala Cys Ala Asn Gln Met Phe Phe Ser			
95	100	105	
Phe Met Phe Gly Phe Thr His Ser Phe Leu Leu Leu Val Met Gly			
110	115	120	
Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Asn Val			
125	130	135	
Leu Met Ser Pro Arg Asp Cys Ala His Leu Val Ala Cys Thr Trp			

140	145	150
Ala Gly Gly Ser Val Met Gly Met Met	Val Thr Thr Ile Val	Phe
155	160	165
His Leu Thr Phe Cys Gly Ser Asn Val	Ile His His Phe Phe	Cys
170	175	180
His Val Leu Ser Leu Leu Lys Leu Ala	Cys Glu Asn Lys Thr	Ser
185	190	195
Ser Val Ile Met Gly Val Met Leu Val	Cys Val Thr Ala Leu	Ile
200	205	210
Gly Cys Leu Phe Leu Ile Ile Leu Ser	Tyr Val Phe Ile Val	Ala
215	220	225
Ala Ile Leu Arg Ile Pro Ser Ala Glu	Gly Arg His Lys Thr	Phe
230	235	240
Ser Thr Cys Val Ser His Leu Thr Val	Val Val Thr His Tyr	Ser
245	250	255
Phe Ala Ser Phe Ile Tyr Leu Lys Pro	Lys Gly Leu His Ser	Met
260	265	270
Tyr Ser Asp Ala Leu Met Ala Thr Thr	Tyr Thr Val Phe Thr	Pro
275	280	285
Phe Leu Ser Pro Ile Ile Phe Ser Leu	Arg Asn Lys Glu Leu	Lys
290	295	300
Asn Ala Ile Asn Lys Asn Phe Tyr Arg	Lys Phe Cys Pro Pro	Ser
305	310	315
Ser		

<210> 17
<211> 370
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475165CD1

<400> 17			
Met Leu Val Leu Asn Ser Trp Ala Gln Val	Ile His Trp Pro Gln		
1	5	10	15
Pro Pro Lys Val Leu Gly Leu Gln Pro	Leu Glu Lys Thr Gln Tyr		
20	25	30	
Gly Phe Leu Gly Thr Asp Arg Val Glu	Glu Lys Thr Ser Val Ile		
35	40	45	
Thr Ile Arg Val Ser Val Thr His Arg	His Asn Ser Tyr Met Glu		
50	55	60	
Ala Glu Asn Leu Thr Glu Leu Ser Lys	Phe Leu Leu Leu Gly Leu		
65	70	75	
Ser Asp Asp Pro Glu Leu Gln Pro Val	Leu Phe Gly Leu Phe Leu		
80	85	90	
Ser Met Tyr Leu Val Thr Val Leu Gly	Asn Leu Leu Ile Ile Leu		
95	100	105	
Ala Val Ser Ser Asp Ser His Leu His	Thr Pro Met Tyr Phe Phe		
110	115	120	
Leu Ser Asn Leu Ser Phe Val Asp Ile	Cys Phe Ile Ser Thr Thr		
125	130	135	
Val Pro Lys Met Leu Val Ser Ile Gln	Ala Arg Ser Lys Asp Ile		
140	145	150	
Ser Tyr Met Gly Cys Leu Thr Gln Val	Tyr Phe Leu Met Met Phe		
155	160	165	
Ala Gly Met Asp Thr Phe Leu Leu Ala	Val Met Ala Tyr Asp Arg		
170	175	180	
Phe Val Ala Ile Cys His Pro Leu His	Tyr Thr Val Ile Met Asn		
185	190	195	
Pro Cys Leu Cys Gly Leu Leu Val Leu	Ala Ser Trp Phe Ile Ile		

200	205	210
Phe Trp Phe Ser Leu Val His Ile Leu Leu Met Lys Arg Leu Thr		
215	220	225
Phe Ser Thr Gly Thr Glu Ile Pro His Phe Phe Cys Glu Pro Ala		
230	235	240
Gln Val Leu Lys Val Ala Cys Ser Asn Thr Leu Leu Asn Asn Ile		
245	250	255
Val Leu Tyr Val Ala Thr Ala Leu Leu Gly Val Phe Pro Val Ala		
260	265	270
Gly Ile Leu Phe Ser Tyr Ser Gln Ile Val Ser Ser Leu Met Gly		
275	280	285
Met Ser Ser Thr Lys Gly Lys Tyr Lys Ala Phe Ser Thr Cys Gly		
290	295	300
Ser His Leu Cys Val Val Ser Leu Phe Tyr Gly Thr Gly Leu Gly		
305	310	315
Val Tyr Leu Ser Ser Ala Val Thr His Ser Ser Gln Ser Ser Ser		
320	325	330
Thr Ala Ser Val Met Tyr Ala Met Val Thr Pro Met Leu Asn Pro		
335	340	345
Phe Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Gly Ala Leu Glu		
350	355	360
Arg Leu Leu Ser Arg Ala Asp Ser Cys Pro		
365	370	

<210> 18

<211> 318

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7475273CD1

<400> 18

Met Lys Asn Val Thr Glu Val Thr Leu Phe Val Leu Lys Gly Phe			
1	5	10	15
Thr Asp Asn Leu Glu Leu Gln Thr Ile Phe Phe Leu Phe Leu			
20	25	30	
Ala Ile Tyr Leu Phe Thr Leu Met Gly Asn Leu Gly Leu Ile Leu			
35	40	45	
Val Val Ile Arg Asp Ser Gln Leu His Lys Pro Met Tyr Tyr Phe			
50	55	60	
Leu Ser Met Leu Ser Ser Val Asp Ala Cys Tyr Ser Ser Val Ile			
65	70	75	
Thr Pro Asn Met Leu Val Asp Phe Thr Thr Lys Asn Lys Val Ile			
80	85	90	
Ser Phe Leu Gly Cys Val Ala Gln Val Phe Leu Ala Cys Ser Phe			
95	100	105	
Gly Thr Thr Glu Cys Phe Leu Leu Ala Ala Met Ala Tyr Asp Arg			
110	115	120	
Tyr Val Ala Ile Tyr Asn Pro Leu Leu Tyr Ser Val Ser Met Ser			
125	130	135	
Pro Arg Val Tyr Met Pro Leu Ile Asn Ala Ser Tyr Val Ala Gly			
140	145	150	
Ile Leu His Ala Thr Ile His Thr Val Ala Thr Phe Ser Leu Ser			
155	160	165	
Phe Cys Gly Ala Asn Glu Ile Arg Arg Val Phe Cys Asp Ile Pro			
170	175	180	
Pro Leu Leu Ala Ile Ser Tyr Ser Asp Thr His Thr Asn Gln Leu			
185	190	195	
Leu Leu Phe Tyr Phe Val Gly Ser Ile Glu Leu Val Thr Ile Leu			
200	205	210	
Ile Val Leu Ile Ser Tyr Gly Leu Ile Leu Leu Ala Ile Leu Lys			

215	220	225
Met Tyr Ser Ala Glu Gly Arg Arg Lys	Val Phe Ser Thr Cys	Gly
230	235	240
Ala His Leu Thr Gly Val Ser Ile Tyr	Tyr Gly Thr Ile Leu	Phe
245	250	255
Met Tyr Val Arg Pro Ser Ser Ser Tyr	Ala Ser Asp His Asp	Met
260	265	270
Ile Val Ser Ile Phe Tyr Thr Ile Val	Ile Pro Leu Leu Asn	Pro
275	280	285
Val Ile Tyr Ser Leu Arg Asn Lys Asp	Val Lys Asp Ser Met	Lys
290	295	300
Lys Met Phe Gly Lys Asn Gln Val Ile	Asn Lys Val Tyr Phe	His
305	310	315

Thr Lys Lys

<210> 19
<211> 321
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476077CD1

<400> 19
Met Glu Ser Pro Asn His Thr Asp Val Asp Pro Ser Val Phe Phe
1 5 10 15
Leu Leu Gly Ile Pro Gly Leu Glu Gln Phe His Leu Trp Leu Ser
20 25 30
Leu Pro Val Cys Gly Leu Gly Thr Ala Thr Ile Val Gly Asn Ile
35 40 45
Thr Ile Leu Val Val Ala Thr Glu Pro Val Leu His Lys Pro
50 55 60
Val Tyr Leu Phe Leu Cys Met Leu Ser Thr Ile Asp Leu Ala Ala
65 70 75
Ser Val Ser Thr Val Pro Lys Leu Leu Ala Ile Phe Trp Cys Gly
80 85 90
Ala Gly His Ile Ser Ala Ser Ala Cys Leu Ala Gln Met Phe Phe
95 100 105
Ile His Ala Phe Cys Met Met Glu Ser Thr Val Leu Leu Ala Met
110 115 120
Ala Phe Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Ala
125 130 135
Thr Ile Leu Thr Asp Thr Ile Ile Ala His Ile Gly Val Ala Ala
140 145 150
Val Val Arg Gly Ser Leu Leu Met Leu Pro Cys Pro Phe Leu Ile
155 160 165
Gly Arg Leu Asn Phe Cys Gln Ser His Val Ile Leu His Thr Tyr
170 175 180
Cys Glu His Met Ala Val Val Lys Leu Ala Cys Gly Asp Thr Arg
185 190 195
Pro Asn Arg Val Tyr Gly Leu Thr Ala Ala Leu Leu Val Ile Gly
200 205 210
Val Asp Leu Phe Cys Ile Gly Leu Ser Tyr Ala Leu Ser Ala Gln
215 220 225
Ala Val Leu Arg Leu Ser Ser His Glu Ala Arg Ser Lys Ala Leu
230 235 240
Gly Thr Cys Gly Ser His Val Cys Val Ile Leu Ile Ser Tyr Thr
245 250 255
Pro Ala Leu Phe Ser Phe Phe Thr His Arg Phe Gly His His Val
260 265 270
Pro Val His Ile His Ile Leu Leu Ala Asn Val Tyr Leu Leu Leu

275	280	285
Pro Pro Ala Leu Asn Pro Val Val Tyr	Gly Val Lys Thr Lys	Gln
290	295	300
Ile Arg Lys Arg Val Val Arg Val Phe	Gln Ser Gly Gln Gly	Met
305	310	315
Gly Ile Lys Ala Ser Glu		
320		

<210> 20

<211> 313

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7476113CD1

<400> 20

Met Leu Leu Thr Asp Arg Asn Thr Ser Gly Thr Thr Phe Thr Leu			
1	5	10	15
Leu Gly Phe Ser Asp Tyr Pro Glu Leu Gln Val Pro Leu Phe Leu			
20	25	30	
Val Phe Leu Ala Ile Tyr Asn Val Thr Val Leu Gly Asn Ile Gly			
35	40	45	
Leu Ile Val Ile Ile Lys Ile Asn Pro Lys Leu His Thr Pro Met			
50	55	60	
Tyr Phe Phe Leu Ser Gln Leu Ser Phe Val Asp Phe Cys Tyr Ser			
65	70	75	
Ser Ile Ile Ala Pro Lys Met Leu Val Asn Leu Val Val Lys Asp			
80	85	90	
Arg Thr Ile Ser Phe Leu Gly Cys Val Val Gln Phe Phe Phe			
95	100	105	
Cys Thr Phe Val Val Thr Glu Ser Phe Leu Leu Ala Val Met Ala			
110	115	120	
Tyr Asp Arg Phe Val Ala Ile Cys Asn Pro Leu Leu Tyr Thr Val			
125	130	135	
Asn Met Ser Gln Lys Leu Cys Val Leu Leu Val Val Gly Ser Tyr			
140	145	150	
Ala Trp Gly Val Ser Cys Ser Leu Glu Leu Thr Cys Ser Ala Leu			
155	160	165	
Lys Leu Cys Phe His Gly Phe Asn Thr Ile Asn His Phe Phe Cys			
170	175	180	
Glu Phe Ser Ser Leu Leu Ser Leu Ser Cys Ser Asp Thr Tyr Ile			
185	190	195	
Asn Gln Trp Leu Leu Phe Phe Leu Ala Thr Phe Asn Glu Ile Ser			
200	205	210	
Thr Leu Leu Ile Val Leu Thr Ser Tyr Ala Phe Ile Val Val Thr			
215	220	225	
Ile Leu Lys Met Arg Ser Val Ser Gly Arg Arg Lys Ala Phe Ser			
230	235	240	
Thr Cys Ala Ser His Leu Thr Ala Ile Thr Ile Phe His Gly Thr			
245	250	255	
Ile Leu Phe Leu Tyr Cys Val Pro Asn Ser Lys Asn Ser Arg His			
260	265	270	
Thr Val Lys Val Ala Ser Val Phe Tyr Thr Val Val Ile Pro Met			
275	280	285	
Leu Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Asp			
290	295	300	
Thr Val Thr Glu Ile Leu Asp Thr Lys Val Phe Ser Tyr			
305	310		

<210> 21

<211> 328

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476117CD1

<400> 21

Met	Phe	Leu	Thr	Glu	Arg	Asn	Thr	Thr	Ser	Glu	Ala	Thr	Phe	Thr
1									10					15
Leu	Leu	Gly	Phe	Ser	Asp	Tyr	Leu	Glu	Leu	Gln	Ile	Pro	Leu	Phe
									20					30
Phe	Val	Phe	Leu	Ala	Val	Tyr	Gly	Phe	Ser	Val	Val	Gly	Asn	Leu
									35					45
Gly	Met	Ile	Val	Ile	Ile	Lys	Ile	Asn	Pro	Lys	Leu	His	Thr	Pro
									50					60
Met	Tyr	Phe	Phe	Leu	Asn	His	Leu	Ser	Phe	Val	Asp	Phe	Cys	Tyr
									65					75
Ser	Ser	Ile	Ile	Ala	Pro	Met	Met	Leu	Val	Asn	Leu	Val	Val	Glu
									80					90
Asp	Arg	Thr	Ile	Ser	Phe	Ser	Gly	Cys	Leu	Val	Gln	Phe	Phe	Phe
									95					105
Phe	Cys	Thr	Phe	Val	Val	Thr	Glu	Leu	Ile	Leu	Phe	Ala	Val	Met
									110					120
Ala	Tyr	Asp	His	Phe	Val	Ala	Ile	Cys	Asn	Pro	Leu	Leu	Tyr	Thr
									125					135
Val	Ala	Ile	Ser	Gln	Lys	Leu	Cys	Ala	Met	Leu	Val	Val	Val	Leu
									140					150
Tyr	Ala	Trp	Gly	Val	Ala	Cys	Ser	Leu	Thr	Leu	Ala	Cys	Ser	Ala
									155					165
Leu	Lys	Leu	Ser	Phe	His	Gly	Phe	Asn	Thr	Ile	Asn	His	Phe	Phe
									170					180
Cys	Glu	Leu	Ser	Ser	Leu	Ile	Ser	Leu	Ser	Tyr	Pro	Asp	Ser	Tyr
									185					195
Leu	Ser	Gln	Leu	Leu	Leu	Phe	Thr	Val	Ala	Thr	Phe	Asn	Glu	Ile
									200					210
Ser	Thr	Leu	Leu	Ile	Ile	Leu	Thr	Ser	Tyr	Ala	Phe	Ile	Ile	Val
									215					225
Thr	Thr	Leu	Lys	Met	Pro	Ser	Ala	Ser	Gly	His	Arg	Lys	Val	Phe
									230					240
Ser	Thr	Cys	Ala	Ser	His	Leu	Thr	Ala	Ile	Thr	Ile	Phe	His	Gly
									245					255
Thr	Ile	Leu	Phe	Leu	Tyr	Cys	Val	Pro	Asn	Ser	Lys	Asn	Ser	Arg
									260					270
His	Thr	Val	Lys	Val	Ala	Ser	Val	Phe	Tyr	Thr	Val	Val	Ile	Pro
									275					285
Leu	Leu	Asn	Pro	Leu	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Asp	Val	Lys
									290					300
Asp	Ala	Ile	Arg	Lys	Ile	Ile	Asn	Thr	Lys	Tyr	Phe	His	Ile	Lys
									305					315
His	Arg	His	Trp	Tyr	Pro	Phe	Asn	Phe	Val	Ile	Glu	Gln		
									320					325

<210> 22
<211> 324
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476079CD1

<400> 22

Met	Asn	His	Met	Ser	Ala	Ser	Leu	Lys	Ile	Ser	Asn	Ser	Ser	Lys
1									10					15
Phe	Gln	Val	Ser	Glu	Phe	Ile	Leu	Leu	Gly	Phe	Pro	Gly	Ile	His
						20			25					30
Ser	Trp	Gln	His	Trp	Leu	Ser	Leu	Pro	Leu	Ala	Leu	Leu	Tyr	Leu
						35			40					45
Ser	Ala	Leu	Ala	Ala	Asn	Thr	Leu	Ile	Leu	Ile	Ile	Ile	Trp	Gln
						50			55					60
Asn	Pro	Ser	Leu	Gln	Gln	Pro	Met	Tyr	Ile	Phe	Leu	Gly	Ile	Leu
						65			70					75
Cys	Met	Val	Asp	Met	Gly	Leu	Ala	Thr	Ile	Ile	Pro	Lys	Ile	
						80			85					90
Leu	Ala	Ile	Phe	Trp	Phe	Asp	Ala	Lys	Val	Ile	Ser	Leu	Pro	Glu
						95			100					105
Cys	Phe	Ala	Gln	Ile	Tyr	Ala	Ile	His	Phe	Phe	Val	Gly	Met	Glu
						110			115					120
Ser	Gly	Ile	Leu	Leu	Cys	Met	Ala	Phe	Asp	Arg	Tyr	Val	Ala	Ile
						125			130					135
Cys	His	Pro	Leu	Arg	Tyr	Pro	Ser	Ile	Val	Thr	Ser	Ser	Leu	Ile
						140			145					150
Leu	Lys	Ala	Thr	Leu	Phe	Met	Val	Leu	Arg	Asn	Gly	Leu	Phe	Val
						155			160					165
Thr	Pro	Val	Pro	Val	Leu	Ala	Ala	Gln	Arg	Asp	Tyr	Cys	Ser	Lys
						170			175					180
Asn	Glu	Ile	Glu	His	Cys	Leu	Cys	Ser	Asn	Leu	Gly	Val	Thr	Ser
						185			190					195
Leu	Ala	Cys	Asp	Asp	Arg	Arg	Pro	Asn	Ser	Ile	Cys	Gln	Leu	Val
						200			205					210
Leu	Ala	Trp	Leu	Gly	Met	Gly	Ser	Asp	Leu	Ser	Leu	Ile	Ile	Leu
						215			220					225
Ser	Tyr	Ile	Leu	Ile	Leu	Tyr	Ser	Val	Leu	Arg	Leu	Asn	Ser	Ala
						230			235					240
Glu	Ala	Ala	Ala	Lys	Ala	Leu	Ser	Thr	Cys	Ser	Ser	His	Leu	Thr
						245			250					255
Leu	Ile	Leu	Phe	Phe	Tyr	Thr	Ile	Val	Val	Val	Ile	Ser	Val	Thr
						260			265					270
His	Leu	Thr	Glu	Met	Lys	Ala	Thr	Leu	Ile	Pro	Val	Leu	Leu	Asn
						275			280					285
Val	Leu	His	Asn	Ile	Ile	Pro	Pro	Ser	Leu	Asn	Pro	Thr	Val	Tyr
						290			295					300
Ala	Leu	Gln	Thr	Lys	Glu	Leu	Arg	Ala	Ala	Phe	Gln	Lys	Val	Leu
						305			310					315
Phe	Ala	Leu	Thr	Lys	Glu	Ile	Arg	Ser						
						320								

<210> 23
<211> 315
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476112CD1

<400> 23
Met Gln Gly Leu Asn His Thr Ser Val Ser Glu Phe Ile Leu Val
1 5 10 15
Gly Phe Ser Ala Phe Pro His Leu Gln Leu Met Leu Phe Leu Leu
20 25 30
Phe Leu Leu Met Tyr Leu Phe Thr Leu Leu Gly Asn Leu Leu Ile
35 40 45
Met Ala Thr Val Trp Ser Glu Arg Ser Leu His Met Pro Met Tyr
50 55 60

Leu Phe Leu Cys Ala Leu Ser Ile Thr Glu Ile Leu Tyr Thr Val
 65 70 75
 Ala Ile Ile Pro Arg Met Leu Ala Asp Leu Leu Ser Thr Gln Arg
 80 85 90
 Ser Ile Ala Phe Leu Ala Cys Ala Ser Gln Met Phe Phe Ser Phe
 95 100 105
 Ser Phe Gly Phe Thr His Ser Phe Leu Leu Thr Val Met Gly Tyr
 110 115 120
 Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Asn Val Leu
 125 130 135
 Met Ser Leu Arg Gly Cys Thr Cys Arg Val Gly Cys Ser Trp Ala
 140 145 150
 Gly Gly Leu Val Met Gly Met Val Val Thr Ser Ala Ile Phe His
 155 160 165
 Leu Ala Phe Cys Gly His Lys Glu Ile His His Phe Phe Cys His
 170 175 180
 Val Pro Pro Leu Leu Lys Leu Ala Cys Gly Asp Asp Val Leu Val
 185 190 195
 Val Ala Lys Gly Val Gly Leu Val Cys Ile Thr Ala Leu Leu Gly
 200 205 210
 Cys Phe Leu Leu Ile Leu Leu Ser Tyr Ala Phe Ile Val Ala Ala
 215 220 225
 Ile Leu Lys Ile Pro Ser Ala Glu Gly Arg Asn Lys Ala Phe Ser
 230 235 240
 Thr Cys Ala Ser His Leu Thr Val Val Val Val His Tyr Gly Phe
 245 250 255
 Ala Ser Val Ile Tyr Leu Lys Pro Lys Gly Pro Gln Ser Pro Glu
 260 265 270
 Gly Asp Thr Leu Met Gly Ile Thr Tyr Thr Val Leu Thr Pro Phe
 275 280 285
 Leu Ser Pro Ile Ile Phe Ser Leu Arg Asn Lys Glu Leu Lys Val
 290 295 300
 Ala Met Lys Lys Thr Cys Phe Thr Lys Leu Phe Pro Gln Asn Cys
 305 310 315

<210> 24
 <211> 2739
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475208CB1

<400> 24
 atgctgggcc ctgctgtcct gggcctcagc ctctgggctc tcctgcaccc tgggacgggg 60
 gccccattgt gcctgtcaca gcaacttagg atgaaggggg actacgtgct gggggggctg 120
 ttccccctgg gcgaggccga ggaggctggc ctccgcagcc ggacacggcc cagcagccct 180
 gtgtgcacca ggttcttcctc aaacccgcctg ctctggcac tggccatgaa aatggccgtg 240
 gaggagatca acaaacaagtggatctgcg cccgggctgc gcctgggcta cgaccccttt 300
 gatacgtgct cggagccctgt ggtggccatg aagcccagcc tcatgttccct gccaaggca 360
 ggcagcccg acatcgccgc ctactgcaac tacacgcagt accagccccg tggcgtggct 420
 gtcatcgccgc cccactcgctc agagctgccc atggtcaccg gcaagttttt cagtttttc 480
 ctcatgcccc aggtggcgcc ccccaccatc acccaccatc acccagccct gcccgtggga 540
 gcccctgtgt caggagatgc ctcttgccct ttgcagggtca gctacgggtc tagcatggag 600
 ctgctgagcg cccgggagac cttccctcc ttcttccgca ccgtgcccag cgaccgtgt 660
 cagctgacgg ccggccgcgg a gctgctgcag gagttcggt ggaactgggt gggcgccctg 720
 ggcagcgcacg acgagtacgcc cccgcaggcc ctgagcatct tctcgccct ggtcggcacc 780
 gcggcatctg catcgccgac gagggcctgg tgccgctgcc ccgtgcaggaa cgtcctgcac 840
 caggtgaacc agagcagcgt gcaggtgggt ctgctgttcg cctccgtgca cggcccccac 900
 gccccttca actacagcat cagcagcagg ctctcgccca aggtgtgggt gggcagcgcag 960
 gcctggctga cctctgaccc ggtcatgggg ctgccccggca tggcccaagat gggcacgggtg 1020

cttggctcc tccagagggg tgcccagctg cacgagttcc cccagtacgt gaagacgcac 1080
 ctggccctgg ccaccgaccc ggcctctgc tctgcctgg gcgagagggg gcagggtctg 1140
 gaggaggacg tgggggccca gcgctccccg cagtgtact gcatcacgct gcagaaccgt 1200
 gcccaggccc tgacacaacac tcttcagtg aacgcctcag gctgcccgc gcaggacccc 1260
 gtgaagccct ggtagctcct ggagaacatg tacaacctga cttccacgt gggcgggctg 1320
 ccgctcgccgt tcgacagcag cgaaacgtg gacatggagt acgacctgaa gctgtgggtg 1380
 tggcaggcgt cagtccccag gctccacgac gtgggcaggt tcaacggcag cctcaggaca 1440
 gagegcetga agatccgctg gcacacgtct gacaaccago cgagcagagc cagaccccag 1500
 gcctgtgcgc agaagccctg tgcggcgtg gccaggaggg ccaggtgcgc 1560
 cgggtcaagg ggttccactc ctgctgtac gactgtgtgg actgcgaggc gggcagctac 1620
 cggcaaaaacc cagacgacat cgcctgcacc ttttgtggcc aggatgatgt gtccccggag 1680
 cgaagcacac gctgttcccg ccgcaggctc cggttccctgg catggggcga gccggctgtg 1740
 ctgctgtcgc tcctgtctg gaggctggcg ctgggccttg tgctggctgc ttggggctg 1800
 ttcgttccacc atcgggacag cccactgggtt caggcctcgg gggggccctt ggcctgctt 1860
 ggcctgggtg gcctgggcct ggtctgcctc agcgtcctcc tttccctgg ccagcccgac 1920
 cctgcccgt gcctggccca gcagcccttgc tcccacctcc cgctcacggg ctgcctgagc 1980
 acactttcc tgcaggccgc cgagatcttgc tggagtcag aactgcctct gagctggca 2040
 gaccggctga gtggctgcct gcggggggcc tgggcctggc tgggtggct gctggccatg 2100
 ctggtggagg tcgactgtg cacctggta ctgggtggct tccgcggga gttgggtgact 2160
 ggactggcac atgtgcccc cggaggcgct ggtgactgc cgcacacgtc ctgggtcag 2220
 cttcggcta gcgcacgcca ccatgcacgc ctggccttgc tctgcttcac tggcacttt 2280
 cctggtgcgg agccagccgg gccgctacaa ccgtgcacgc tggcctcaca catttgccat 2340
 gctggcctac ttcatcacac tgggtctcact tttgtggccc tccggcaca atgtgcagg 2400
 ggtcaactcag gcccagccgt gcagatgggc gcctccctgc tctgtgtcct gggcatactg 2460
 gctgccttcc acctgcccag gtgttacctg ctcatgcggc agccagggtc caacaccccc 2520
 gagttttcc tgggaggggg ccctgggat gccacaaggc cagaatgacg ggaacacacagg 2580
 aaatcaggggg aaacatgggt gacccaacca ctgtatctc agcccccgtg aacccagact 2640
 tagtgcgat ccccccggcc agcaatga cccgtgtctc gctacagaga ccctcccgct 2700
 ctaggttctg accccaggtt gtcctgtac ctgacccccc 2739

<210> 25
 <211> 993
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475101CB1

<400> 25
 atggaagggtt tttatctgcg cagatcacac gaactacaag ggatggaaa accaggcaga 60
 gtgaaccaaa ccactgttcc agacttctc ctctaggac tctctgagtg gccagaggag 120
 cagcctcttc tggggcat ctcccttggc atgtacctgg tcaccatggt ggggaacctg 180
 ctcattatcc tggccatcag ctctgaccca cacctccata ctcccatgtt ctctttctg 240
 gccaacctgt cattaactga tgcctgttgc acttctgcct ccatccccaa aatgctggcc 300
 aacattcata cccagagtca gatcatctcg tattctgggt gtcttgcaca gctatatttc 360
 ctcccttatgt ttgggtggct tgacaactgc ctgctggctg tgatggcata tgaccgctat 420
 gtggccatct gccaaccact ccattacagc acatctatga gtcccccagct ctgtgcacta 480
 atgctgggtg tggctgggt gctaaccacat tgcctgtccc tgatgcacac actgttgcgt 540
 accccgcgtgg ctttctgtgc ccagaaagcc atccctcatt tctattgtga tcttagtgc 600
 ctccctgaagg ttgcctgtc agataccat gtaaaacgagc tgatgtatcat caccatggc 660
 ttgctgttcc tcaactgttcc ctcctgtctg atcgtcttct cctatgtccg cattttctgg 720
 gctgtgttttgc tcatctcatt tcctggaggg agatggaaagg ctttctctac ctgtgggtct 780
 catctcactgg tgggtctgtct cttctatggg tctcttatgg gtgtgtatattt attcctcca 840
 tcaacttacttactacagagag ggaaagtagg gctgctgttc tctatatggt gattattccc 900
 acgctaaacc cattcattta tagcttgagg aacagagaca tgaaggagggc tttgggtaaa 960
 ctttttgtca gtggaaaaac attctttta tga 993

<210> 26
 <211> 990
 <212> DNA
 <213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475152CB1

<220>
<221> unsure
<222> 1
<223> a, t, c, g, or other

<400> 26
ngtgagtaca agtccatggg aatgtccaac ctgacaagac tctctgaatt tattctcttg 60
ggactctcct ctccgtctga agaccaggagg ccactctttg ccctcttct tatcatatac 120
ctggtaactt tgatggaaa tctgctcatc atcttggcta tccactctga tcctcgactt 180
caaaaccta tgcattttt cctaagcatc ttgcctttg ctgatatttg ctacacaaca 240
gtcatagtcc caaagatgct cgtgaacttc ttatcagaga aaaagaccat ttcttatgct 300
gaatgtctgg cacagatgta ttcttcctg gttttggaa acatagatag ttatctcttg 360
gcggctatgg ccatcaaccg ctgttagcc attttaacc cattccatta tgtcactgtt 420
atgaaccgca gatgctgtgt gttgctacta gcattccca tcacttctc ctatttccac 480
tctctctac atgtctctt ggtgaatcgg ctaccccttt gtacatcaa ttttatccat 540
cattttttt gtatgtcaa ccctgtctg aaactgtcct gtcctccac ctttgtcaat 600
gaaattgtgg ccatgacaga agggctggcc tctgtatgg ctccatttgc ctgtatcatc 660
atctcttatac taagaattct catcgctgtt ctcacatcc cctcaggcagc tgaaaacac 720
aaagccttct ccacctgcag ctcccatctc actgtggta ttctgtttt tgaggatt 780
agctatgtct attgcagcc ttgtccagc tatactgtca aggaccaat agcaacaatc 840
aactacactg tggacatc agtggtaac ccatttatct acagtttaag aaacaaagac 900
atgaaacggg gcttacagaa attgataaac aagattaatgtagtcaatgtagtctct 960
acaaaagacca ataaaatctg tggaccctga 990

<210> 27
<211> 1125
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475164CB1

<400> 27
atggccatct gtaacccgct tctgtataac attgccatgt cccctaaatgt gtgttccagc 60
catatgttg gttccactt ctggcccttt tctggggcca tggcccatac cagggtgcatt 120
ctgaaactga cctccgtga ggcacacacc atcaaccact acttctgtga cacccttcat 180
ctgctccagc tctcttgcac cagcacctac gtcagggtctg agtttatctt ggaggcttg 240
acacaacgccc cagaacttca actgcccactc ttccctctgt tccttggaaat atatgtggtc 300
acagtggtgg ggaacctggg catgatctt ttaattgtca tcagttctca actttaccct 360
ccagtgtatt atttctcag tcatttgttca ttcatgttgc tctgtactc ctctgtcatt 420
acccctaaaga tgctggtaa ctttgttca gaggagaaca ttatctctt tctggatgc 480
attactcaac ttatatttctt ctttatattt gtaattgtcag aaggctatct tctgacagcc 540
atggaatatg accgttatgt tgctatctgt cgcccaactgc ttacaatat tgcattgtcc 600
cacagggtct gttccataat gatggctgtg gtaacttcaat tgggtttctt gtggccaca 660
gtccatacta cccgcattgtc agtgggtca ttctgttaggt ctcatacggt cagtcattat 720
ttttgtgata ttcccccattt attgacttgc tcttgcattca gcacccacat caatgagatt 780
ctgctttca ttattggagg agttaatacc tttagcaacta cactggcggt ctttatctct 840
tatgtttca ttttctttagt tttcttgcattca ctgaggggca atccaaagcc 900
tttggcactt gtagctccca tctcttgcatt gtggccatct ttttgggtc tataacattc 960
atgtatttca agcccccttc cagcaacttca atggaaaaag agaagggttc ttctgtgttc 1020
tacatcacaa taatccccat gctgaatctt ctaatctata gcctgaggaa caaggatgtg 1080
aaaaatgcac tgaagaatgacttagggaa aggcagtcat cctga 1125

<210> 28
<211> 939
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475170CB1

<400> 28
atggatcaga aaaatggaag ttcttcact ggatttatcc tactgggtt ctctgacagg 60
cctcagctgg agctagtcct ctttgtggtt cttttgatct tctatatctt cactttgctg 120
gggaacaaaa ccacatcattgt attatctcac ttggaccac acatccacac tcctatgtat 180
ttttttctt ccaacctaag cttttggat ctgtgttaca caacccgcata tggtccacag 240
ctctggta atctcagggg agcagacaaa tcaatctcct atgggtgtt tgtagttcag 300
ctgtacatct ctctaggctt gggatctaca gaatgcgttc tcttaggagt gatggattt 360
gaccgctatg cagctgttg caggcccctc cactacacag tagtcatgca cccttgc 420
tatgtgtga tggcttctac ttcatgggtc attgggtttt ccaactccctt attgcagacg 480
gtgctcatct tgcttttaac actttgtgaa agaaataaat tagaacactt tctttgtgag 540
gttcctccat tgctcaagct tgctgtgtt gacactacta tgaatgaatc tgaactctc 600
tttgcgttg tcattattct tcttgatctt gttgcattaa tcatattctc ctatagtcag 660
attgtcaggc cagtcatgag gataaagtta gcaacaggc agagaaaagt gtttgggaca 720
tgtggctccc acetcacagt ggtttccctg ttctacggca cagctatcta tgcttacctc 780
cagccggca acaactactc tcaggatcag gcgaagttca tctctctt ctacaccatc 840
attacaccca tgatcaaccc cctcatatat acactgagga acaaggatgt gaaaggagca 900
cttaagaagg tgctctggaa gaactacgac tccagatga 939

<210> 29
<211> 978
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475197CB1

<400> 29
atgaagactt ttagttccctt tttcagatc ggcagaaata tgcataagg aaaccaaacc 60
accatcaactg aattcattct cctgggattt ttcaagcagg atgagcatca aaaccccttc 120
tttgtgtttt tttgggtat gtacctggtc actgtgattt ggaacggctt catcattgtg 180
gctatcagct tggatacgtt ctttcatacc cccatgtatc tcttccttgc caatctatcc 240
tttgcgtata ttccctccat ttccaaactca gtccccaaa tgctgtgaa tattcaaacc 300
aagagtcaat ccatctctta tgagagctgc atcacacaga tgtactttt tattgtgttt 360
gtcgctattt acaatttgctt cttggggacc atggcctatg accactttgtt ggcgatctgc 420
caccctctga attatacaat tctcatgcgg cccaggttcg gcattttgtt cacagtcatc 480
tcatgggttcc tcaagtaatat tattgtctgc acacacaccc ttctgctcat tcaattgtctc 540
ttctgttaacc acaacactctt cccacacttc ttctgtactt tggccctctt gctcaaactg 600
tcctgttcag atacattgtt caatgagctt gtgtgtttt ttgtgggtttt atcagttatc 660
atcttccctt ttacacttcag cttctttcc tatgtctgca tcatcagagc tgccttgaga 720
gtatcttcca cacagggaaa gtggaaagcc ttctccactt gtggctctca cctgacagt 780
gttattactgt tctacggaa cattttaggc gtgtactttt tcccttcctc cactcacct 840
gaggacactg ataagattgg tgctgtccta ttcaactgtgg tgacacccat gataaaacccc 900
ttcatctaca gcttgaggaa taaggatatg aaaggtgcc tggaaaagct catcaataga 960
aaaatttctt ccctttga 978

<210> 30
<211> 936
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475210CB1

<400> 30
atggaaaacc aatccagcat ttctgaattt ttctccgag gaatatcagc' gcctccagag 60
caacagcagt ccctttcgg aattttccctg tttatgtatc ttgtcacctt gactggaaac 120
ctgctcatca ttctggccat tggctctgac ctgcacccctt acaccccat gtacttttc 180

ttggccaacc tgtctttgt tgacatgggt ttaacgtccct ccacagttac caagatgctg 240
 gtgaatatac agactcgca tcacaccatc tcctatacgg gttgcctcac gcaaattgtat 300
 ttctttctga tgggggtgt tcttagacagg ttcttcctgg ctgccatggc gtatgaccgc 360
 tatgtggcca ttggccacc cctctgtac tccacagtca tgagggccccca agtctgtgcc 420
 ctaatgcttg cattgtgctg ggtcctcacc aatatcggtt ccctgactca cacgttccctc 480
 atggctcggt tgccttctgt tggactggg gaaattgctc acttttctg tgacatcact 540
 cctgtcctga agctgtcatg ttctgacacc cacatcaacg agatgatggg ttttgtctg 600
 ggaggcaccg tactcatcg cccctttta tgcattgtca cctccatcat ccacattgtg 660
 ccagctatcc tgaggggtccg aaccgtggt ggggtggca aggcccttcc cacctgcagt 720
 tcccacctct gcgttgcgtt tgcgttctat gggaccctct tcagtgcccta cctgtgtcc 780
 ccctccatttgc cctctgaaga gaaggacatt gcagcagctg caatgtacac catagtgact 840
 cccatgttga acccccttat ctatagccata aggaacaagg acatgaaggg ggcctaaag 900
 aggctttca gtcacaggag tattgttcc tcttag 936

<210> 31
 <211> 1035
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475221CB1

<400> 31
 atggagcttc tgacaaataa tctcaaattt atcaactgacc cttttgtttt taggctccga 60
 cacctgagtc caacacatcc agaagaacac atgaaaaata agaacaatgt gactgaattt 120
 atcctttag ggctcacaca gaacctgag gggcaaaaagg ttttattttt cacattctta 180
 ctaatctaca tggtgacgat aatggcaac ctgcttataca tagtaccat catggccagc 240
 cagtcctgg gttcccccattt gtactttttt ctggcttctt tatttttccat agataccgtc 300
 tattttactg catttgcgtt caaaatgattt gttgacttgc tctctgagaa aaagaccatt 360
 tccttcagg gttgtatggc tcaactttttt atggatcatt tattttgttgg tgctgaagtc 420
 attcttcgg tggtaatggc ctatgatcga tacatggcca tctgttaagcc tcttcatgaa 480
 ttgatcacca tgaatcgctg agtctgtgtt cttatgtgtt tggccgcctg gattggaggc 540
 tttcttcact cattggttca attttctttt attttatcagc tccctttctg tggacccaat 600
 gtcattgaca acttcctgtg tgattttgtat cccttatttga aacttgcctt caccaataacc 660
 tatgtactg ggctttctat gatagctaat ggaggagcga tttgtgttgcgtt caccttcttc 720
 actatccctgc tttccatattt ggtcatattt cactcttta agactcagag tttggaaagg 780
 aaacgaaaag ctttctacac ctgtgcattcc cacgtcactg tggtcattttt attctttgtc 840
 ccctgtatct ttttgcgtt aaggcccaat tttacttttccatgttccatgataa atccatgact 900
 gtagttctaa ttttataac tccatgttgc aacccacttta ttttgcgttccatgataa 960
 gaaatgaaaaa gtgcgttgc gaaactttgg agtaaaaaaag taagcttagc tggaaatgg 1020
 ctgtatcact catga 1035

<210> 32
 <211> 942
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475244CB1

<400> 32
 atggcatctg aaagaaaatca aagcagcaca cccacttttta ttctcttggg tttttcagaa 60
 taccctggaaa tccaggttcc actctttctg gttttcttgc tgcgttctacac agtcaactgt 120
 gtggggaaact tgggcgtat aataatcatc agactcaattt caaaactccca tacaatcatg 180
 tacttttcc tttagtcaactt gtccttgcata gacttctgtt tttccactgtt agttacacct 240
 aaactgttgg agaacttggt tggaaatac agaaccatctt cttttcttgc tgcgttcatg 300
 caattttgtt ttgcttgcattt ttttggagtg acagaaaactt tcatgtttagc agcgtatggct 360
 tatgaccgtt ttgtggcagt ttgttgcattt ttttgcgttata ccactattat gtctcagaag 420
 ctctgtgtcc ttctgggtgc tgggtcctat acatggggga tagtgcgttcc cctgtatctc 480
 acatattttc ttcttgactt atcgttttgcgtt gatctactt ttttgcgttata ttgttcatgt 540
 gaccactctg taattttttc tgccttgcattt ttttgcgttccatgataa atatcagccca gaggctatgc 600

tttatttattg ccatattcaa tgaggtgagc agcctaatta tcattctgac atcatatatatg 660
 ctatatttca ctaccattat gaagatgcga tctgcaagt ggcgccagaa aactttctcc 720
 acctgtgcct cccacctgac agccatcaat atcttccatg gaactatcct tttccttac 780
 tgtgttccta atcctaaaac ttctagcctc atagttacag tggcttctgt gtttacaca 840
 gtggcgattc caatgctgaa cccattgatc tacagcctta ggaacaaaga tatcaataac 900
 atgtttaaa aatttagttgt caccaaattt atttaccact ga 942

<210> 33
 <211> 942
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475293CB1

<400> 33
 atgaagaggg agaatcagag cagtgtgtct gagttcctcc tcctggaccc cccatctgg 60
 ccagagcagc aggctgtgtt cttcacccctg ttcttggca tgtacctgat cacgggtctg 120
 gggAACCTGC tcacatccct gtcatecgg ctggactctc accttcacac cccatgttc 180
 ttcttcctca gccacttggc tctcaactgac atetcctt catctgtcac tggccaaag 240
 atgttattaa gcatgcaaac tcaggatcaa tccattctt atgcagggtg tgaactcag 300
 atgttatttt tcacatccctt cactgatcta gacaatttcc ttctcacttc aatggcatac 360
 gatcggtatg tggccatctg tcacccctc cgctacacca ctatcatgaa agaggactg 420
 tgtaacttac tagtcaactgt gtcctggatc ctctcctgta ccaatggccct gtctcacact 480
 ctccctctgg cccagctgtc cttttgtgct gacaacacca tccccatctt ctctgtgat 540
 ctgttgccc tactcaagct tcacatgctca gacatctccc tcaatgagct ggtcattttc 600
 acagtgggac aggcaatgt tactctacca ctaatatgca tcttgatctc ttatggccac 660
 attgggtca ccacatccaa ggctccatct actaaggga tcttcaaaagc ttgtccacc 720
 tggccttc acctctctgt ggtgtctgt tattatggca caattattgg actgttatttt 780
 ctccctcat ccagtgcctc cagtgacaag gacgtatgg cctctgtgat gtacacgggtg 840
 atcaccatc tgctgaatcc ttccattat agcctaagga acaggacat aaaggagcc 900
 ctggagagac tcttcaacag ggcaacatgc ttatctcaat ga 942

<210> 34
 <211> 930
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475297CB1

<400> 34
 atggaaaatc aaaacaatgt gactgaattc attttctgg gtctcacaga gaaacctggag 60
 ctgtggaaaa tattttctgc tggtttctt gtcatgtatg tagccacagt gctggaaaaat 120
 ctacttattt ggttaactat tatcacaagt cagagtctga ggtcacctat gtatTTTTT 180
 cttacccctt tggccctttt ggtatgtatg ttctcatctg tcgttgcggcc caaggtgatt 240
 gtagacaccc tctccaagag cactaccatc tctctcaaag gctgcctcac ccagctgttt 300
 gtggagcatt tctttggatc tggggatc atccctctca ctgtgtatgc ctatgaccgc 360
 tacgtggcca tctgtaaagcc cctgcactac acgatcatca tgagtccacg ggtgtctgc 420
 ctaatggtag gaggggctt ggtgggggaa ttatgcacg caatgataca acttctttc 480
 atgtatcaa taccctctg tggccataat atcatagatc actttatatg tgatttttt 540
 cagttgtga cacttgcctg cacggacacc cacatccctgg gcctcttagt taccctcaac 600
 agtggatga tgggtgtggc catcttctt atcttaattt cgtcctacac ggtcataatc 660
 tgctccctga agtcttacag ctctaaaggc cggcacaaag ccctctctac ctgcagctcc 720
 cacctcacgg tgggtgtatt gttcttgc ccctgtattt tcttgcacat gaggcctgtg 780
 gtcactcacc ccatagacaa ggcaatggct gtgtcagact caatcatcac acccatgtta 840
 aatcccttga tctataact gaggaatgca gaggtgaaaa gtgccatgaa gaaactctgg 900
 ataaaatggg aggcttggc tggaaataa 930

<210> 35
 <211> 942

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475193CB1

<400> 35
atggaaactg caaattacac caaggtgaca gaatttgc tcactggcct atcccagact 60
ccagagggcc aactagtcc ttgttata ttctatcc tctattgtt catcctacca 120
ggaaatacc ttatcatgg caccatcg ctagaccctc atctgaccc tcctatgtat 180
ttccctgtgg ctaatctgg cttccttgat atttggtaact cttccattac agcccctgaa 240
atgtcatag acttctttgt ggagaggaag ataattctt ttgatggatg cattgcacag 300
ctcttcttc tacactttgc tggggcttcg gagatgttct tgctcacagt gatggccctt 360
gacctctaca ctgctatctg ccgaccctc cactatgcta ccatcatgaa tcaacgtctc 420
tgctgtatcc tggtggctct ctcctggagg gggggcttca ttcatcttat catacagggt 480
gctctcattt ttcgacttcc ttctgtggg cccaatgagt tagacagttt cttctgtgac 540
atcacacagg ttgtccggat tgcctgtgcc aacacccctc cagaggagtt agtgcatttc 600
tgttagtagtg gtctgatctc tgggtgtgt ttgattgtct tgtaatgtc ctatgccttc 660
cttctggcct tggtaaagaa actttcaggc tcaggtgaga ataccaacag gccatgtcc 720
acctgtatt cccacattac cattgtggc ctaatgtttt ggcacatccat ctacatttat 780
gctcggccat ttgactcggtt ttcccttagat aaagtgggtt ctgtgtcaa tacttaata 840
ttcccttac gtaatcccat tatttacaca ttgagaaaca aggaagtaaa ggcagccatg 900
aggaagttgg taccatccat tattttgtt aaagagaagt ga 942

<210> 36
<211> 1029
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475213CB1

<400> 36
atgaagagaa agaacttcac agaagtgtca gaattcattt tcttggatt ttctagcttt 60
ggaaagcatc agataaccct ctttgggtt ttccttaactg tctacatccc aactctgggt 120
gctaacatca tcattgtgac tatcatctgc attgaccatc atctccacac tcccatgtat 180
ttcttctaa gcatgctggc tagttcagag acgggttaca cactgtcat tggccacga 240
atgctttga gcctcatttt tcataaccaa cctatctct tggcaggctg tgctacacaa 300
atgttctttt ttgttatctt ggccactaat aattgttcc tgcttactgc aatggggat 360
gaccgctatg tggccatctg cagaccctg agatacactg tcatcatgag caagggacta 420
tggcccccac tgggtgtgg gtccttggc attggcttgc ctatggcact tctccatgtg 480
acagccatgt tcaatttgc gttctgtggc acagtggtag accacccctt ttgtgacatt 540
tacccagtca tggaaacttcc ttgcattgtt accactatca atgagataat aaattatgg 600
gtaagttcat ttgtgatttt tggccatata ggcctgatata ttatcccta tggccctgtc 660
atctttcca tccttcaat tgcctcactg gagggccggg agaagaccc ttccacactgt 720
gtctcccaacc tcactgtggt tatttccac tgggtgtgt cttccattgc ctaccaag 780
ccgaagtcag aaagttcaat agaaaaagac ttgttctct cagtgcgtt caccatcatc 840
actcccttc tgaaccctgt tggttacactg tggagaaaca aggagataca agaatcactc 900
caagctggat taagactact tgggtgtgtt cttgaagatt tcaagtttga aagcttttg 960
gctcccatcc tacctgaact ctctgacact caaatctttt agcttgcgtt gtttagggat 1020
gtggagtag 1029

<210> 37
<211> 933
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475272CB1

<400> 37
atggcagaga tgaacctcac cttgggacc gagttccctcc ttattgcatt cactgaatat 60
cctgaatggg cactccctct cttccttgc ttattatcta tgtatctcat caccgtattg 120
ggaaacttag agatgattat tctgatccctc atggatcacc agctccacgc tc当地atgtat 180
ttccttcgta gtcacctcgc tttcatggac gctgctact catctatcac tgtccccca 240
atgctggcag tgctgctgga gcatggggca gctttatctt acacacgctg tgctgctcag 300
ttctttctgt tcacccctt gggttccatc gactgctacc tcttggccct catggcctat 360
gaccgtact tggtgtgtg ccagccctg ctttatgtca ccatctgac acaggcaggcc 420
cgcttgagt ttgtggctgg ggcttaacgtt gctggctca tcagtgcctt ggtgcggaca 480
gtctcagccct tcactctc cttctgtgga accagtgaga ttgactttat ttctgtgac 540
ctccctccctc tgtaaagtt gacctgtggg gagagctaca ct当地agaagt gctgattatt 600
atgttgcca tttttgtcat ccctgttcc atgggtgtga tcttgggtgc ct当地ctgttt 660
atcatctgtgg cc当地atggg gatccctgct ggaagccagg ccaagacccct ctccacccctgc 720
accccccacc tcactgctgt gtcactctc ttggtaacc tc当地tctcat gtacttgaga 780
ggtaactcag atcagtc当地tcc ggagaagaat cgggtatgtg ctgtgttta cacagaggcc 840
atccccatgt tgaatccccct catctacagc ctgaggaaca aggaagtgaa ggaggccctg 900
agaaaaattc tcaatagagc caagttgtcc taa 933

<210> 38
<211> 948
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475200CB1

<220>
<221> unsure
<222> 1
<223> a, t, c, g, or other

<400> 38
ngcaatactg cacctgcatt ctcagtgacc ttggaaatcta tggacatacc aaaaaatatac 60
acagaatttt tcatgctgg gctctcacag aactcagagg tacagagat tctcttgc 120
gtcttttgc tcatctatgt ggtcacgggt tggcaaca tgctcattgt ggtcaactatc 180
accccccacc ccacgctggc ttccctgtg tatttttcc tggcaacct atccttatt 240
gacacccttt attcttcttc tatggctct aaactcattt ctgactcattt gtatgagggg 300
agaaccatct cttatgatgt ctgatggct cagcttgc gagctcattt ttgggaggt 360
gttgagatca ttctgctcac agtgatggct tatgaccgt atgtgccat ctgtaagccc 420
ctgcacaata ctaccatcat gaccaggcat ctctgtgcca tgcttgtagg ggtggcttgg 480
cttggggct tcctgcattt attgggtcag ctccctctgg tccttgggt gcccttctgt 540
ggcccaatg tcatcaatca ctttgcctgt gacttgtacc ctttgcgttga agttgcctgc 600
accaatacgt atgtcattgg tctgctgggt gttgccaaca gtggttaat ctgcctgtt 660
aacttcctca tgctggctgc ctccctacatt gtcacccctgt actccctttag gtc当地acagt 720
gcagatggga gatgcaaacc cctctccacc tggagccctt acttcatgt tggtgcctt 780
ttcttgcattt cctgtatatt tacttgcattt catccattttt ctacttacc tatagacaaa 840
aatatggcat tattttatgg tatttgcata cctatgttga atccactcat ttataccctg 900
agaaatgaag agttaaaaaa tgccatgaga aagcttttta catggtaa 948

<210> 39
<211> 951
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475121CB1

<400> 39
atgccttagtc agaactatac catcatatct gaatttaacc tctttggctt ctcagcccttc 60
ccccagcacc tcctgcccattt cttgtccctg ctgtacccctc tgatgttctt gttcacattt 120
ctggcaacc ttctcatcat ggccacaatc tggattgaac acagactcca cacaccatg 180

tacctttct tggcacccct ctccgtctct gagattctgt tcactgtgc catcacccct 240
 cgcacatgtgg ctgatctgtt ttcccaccat cattccatca cctttgtggc ttgtgccaac 300
 cagatgttct tctccttcat gtttggcttc actcactctt tccttctctt ggtcatggc 360
 tatgatcgct atgtggccat ctgccaccca ctgcgttaca atgtgtcat gagccccgt 420
 gactgtgccc atcttggc ctgttacctgg gctgggtggct cagtcatggg gatgtatggtg 480
 acaacgatag tttccaccc cactttctgt gggtctaata tgatccacca tttttctgt 540
 catgtgttt ccctcttcaa gttggcctgt gaaaacaaga catcatctgt catcatggg 600
 gtgatgtgg tgggtgtcac agccctgata ggctgttat tcctcatcat cctctccat 660
 gtcttcattt tggctgccc ctggaggatt ccctctggc aaggccggca caagacattt 720
 tctacgtgtg tattccaccc cactgtggtg gtcacgcact atagtttgc ctcccttata 780
 tacctcaagc ccaaggggctt ccattctatg tacagtgcg ccttgatggc caccacctat 840
 actgtttca ccccttctt tagccaaatc attttcagcc taaggaacaa ggagctgaag 900
 aatgccataa ataaaaaactt ttacagaaaa ttctgtccctc caagttccctg a 951

<210> 40
 <211> 1113
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475165CB1

<400> 40
 atgctggct tgaactcctg ggctcaagtg atccactggc ctcagcctcc caaagtgtg 60
 ggattacagc ctttggaaaa aacccagttac ggcttcttag gaacagatcg tggtaaagag 120
 aaaacttcag tgataaccat cagagttatgt gtgacccaca gacacaacag ctacatggaa 180
 gcagaaaaacc ttacagaattt atcaaattt ctccctctgg gactctcaga tgatcctgaa 240
 ctgcagcccg tcctctttgg gctgttcttg tccatgttacc tggtcacggg gctggggaa 300
 ctgctcatca ttctggccgt cagctctgac tcccacctcc acaccccat gtacttcttc 360
 ctctccaaacc tggctttgt tgacatctgt ttcatctcca ccacagcccc caagatgcta 420
 gtgagcatcc aggacacggag caaagacatc tcctacatgg ggtgcctcac tcaggtgtat 480
 ttttaatga tggataact ttctactgg ccgtgtatggc ctatgaccgg 540
 tttgtggcca tctgccaccc actgcactac acggtcatca tgaaccctg cctctgtggc 600
 ctccctggttc tggcatctt gttcatatt ttctgggttcc ccctgggtca tattctactg 660
 atgaagaggt tgaccccttc cacaggact gagattccgc atttctctg tgaaccggct 720
 caggtccctca aggtggcctg ctctaaacacc ctcctcaata acattgtctt gtatgtggc 780
 acggcactgc tgggtgtgtt tcctgttagt gggatcctct tctcctactc tcagattgtc 840
 tcctccctaa tggaatgtc ctccaccaag ggcaagtaca aagcctttc cacctgtgg 900
 ttcacccctt gtgtggtctc ttgttctat ggaacaggac ttgggtctt tctgagttt 960
 gctgtgaccc attttccca gagcagctcc accgcctcag tgatgtacgc catggtcacc 1020
 cccatgctga acccccttcat ctacagccctg aggaacaagg atgtgaaggg ggcctggaa 1080
 agactccctca gcaggccga ctcttgcca tga 1113

<210> 41
 <211> 957
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7475273CB1

<400> 41
 atgaagaatg tcactgaagt taccttattt gtactgaagg gcttcacaga caatcttggaa 60
 ctgcagacta tcttcttctt cctgtttcta gcaatctacc tcttcactct catgggaaat 120
 ttaggactga ttttagtgtt cattaggat tccctactcc acaaacccat gtactatttt 180
 ctgagttatgt tgctttctgt ggatgcctgc tattccctcag ttattacccc aaatatgtta 240
 gtagattttt cggacaaagaa taaaatgtt cattccctt gatgtgttagc acaggtgttt 300
 ctgtgtgtt gtttggaaac cacagaatgc ttctcttgg ctgcaatggc ttatgtacgc 360
 tatgttagcca tctacaaccc tctctgtat tcagttagca tgtcacccag agtctacatg 420
 ccactcatca atgtttccca tggctgccc attttacatg ctactataca tacagtggct 480
 acattttagcc tattccctctg tggagccat gaaatttaggc gtgtctttt tgatatccct 540

cctctccttg ctatttctta ttctgacact cacacaaacc agcttctact cttctacttt 600
 gttggctcta tcgagctggt cactatcctg attgttctga ttccttatgg tttgattctg 660
 ttggccatcc tgaagatgtt ttctgctgaa gggaggagaa aagtcttctc cacatgtgga 720
 gctcacctaa ctggagtgtc aatttattat gggacaatcc tcttcatgtt tgtgagacca 780
 agttccagct atgcttcgga ccatgacatg atagtgtcaa tattttacac cattgtgatt 840
 cccttgcgtt atcccgcat ctacagttt aggaacaaag atgtaaaaga ctcaatgaaa 900
 aaaatgtttt ggaaaaatca gtttatcaat aaagtatatt ttcataactaa aaaataa 957

<210> 42
 <211> 966
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7476077CB1

<400> 42
 atggaatctc ctaatcacac ttagtggac cttctgtct tttccctcctt gggcatcccc 60
 ggtctggAAC aatttcattt gtggctctca ctcctgtgt gtggcttagg cacagccaca 120
 attgtggca atataactat tctgggtt gttgccactg aaccagtctt gcacaaggct 180
 gtgtacccctt ttctgtcat gctctcaacc atcgacttgg ctgcctctgt ctccacagtt 240
 cccaagctac tggatatctt ctgggtgtgaa gccggacata tatctgcctc tgccgcctg 300
 gcacagatgt ttttcattca tgccttcgc atgatggagt ccactgtgt actggccatg 360
 gccttgcattt gctacgtggc catctgcac ccactccgct atgcccacat cctcaactgac 420
 accatcattt cccacatagg ggtggcagct gtatgcgag gtcctctgtt catgtcccc 480
 tgccttcctcc ttattggcg tttgaacttc tgccaaagcc atgtgatcct acacacgtac 540
 tggagcaca tgggtgttgtt gaagctggcc tggagaca ccaggectaa ccgtgtgtat 600
 gggctgacag ctgcactgtt ggtcattggg gttgacttgtt tttgcattgg tctctccat 660
 gcccataatgtt cacaagctgtt ctttcgcctc tcatccatg aagctcggtc caaggcccta 720
 gggacctgtt gttccatgtt ctgtgtcatc ctcatctttt atacaccagg cctttctcc 780
 tttttacac accgctttgg ccattcagttt ccagtcataa ttcacattt tttggccaaat 840
 gtttatctgc ttggccacc tgctcttaat cctgtggat atggagttaa gaccaaacag 900
 atccgtaaaa gagttgtcag ggtgtttcaa agtggcagg gaatggccat caaggcatct 960
 gagtga 966

<210> 43
 <211> 975
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7476113CB1

<220>
 <221> unsure
 <222> 1
 <223> a, t, c, g, or other

<400> 43
 naactaactt tcagattcga agaaacagaa gcgtatgtgc tgactgatag aaatacaagt 60
 gggaccacgt tcacccttggcttc gattacccag aactgcaagt cccacttttc 120
 ctggtttttc tggccatcta caatgtcaat gtgttagggaa atattgggtt gattgtgatc 180
 atcaaaatca accccaaact gcatacccccc atgtactttt tcctcagcca acttccttt 240
 gtggattttt gctatttcctc catcattgtt cccaaagatgt tggtaaccc tgggtcaaa 300
 gacagaacca tttcattttt aggatgcgtt gtacaattttt ttttcttctg tacctttgtg 360
 gtcactgaat cttttttt agctgtgatg gcctatgacc gcttcgtggc catttgcaac 420
 cctctgtctt acacagttaa catgtcccgaa aactctgtcg tggctgtgtt tggggatcc 480
 tatgcctggg gaggctcatg ttccctggaa ctgacgtgtt ctgtttaaa gttatgtttt 540
 catggttca acacaatcaa tcacttctc tggagttt cctcaactact ctcctttct 600
 tgctctgtata cttacatcaa ccagtggctg ctatttttc ttggccaccc taatgaaatc 660
 agcacactac tcacatcttgc acatcttgc ggttacccat cctcaagatg 720

```

cgttcagtca gtgggcgccc caaaggccttc tccacacctgtc cctcccacct gactgccatc 780
accatcttcc atggcaccat cctcttcctt tactgtgtgc ccaactccaa aaactccagg 840
cacacagtca aagtggcctc tgtgtttac accgtggta tccccatgtt gaatcccctg 900
atctacagtc tgagaaaataa agatgtcaag gatacagtca ccggagatact ggacacccaaa 960
gtcttctctt actga

```

<210> 44
<211> 987
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476117CB1

```

<400> 44
atgtttctga cagagagaaa tacgacatct gaggccacat tcactctt gggcttctca 60
gattacctgg aactgc当地 tccc当地tcc tt当地tatttc tggc当地tcta cggctt当地cagt 120
gtggtaggga atctt当地ggat gatagtgatc atcaaaaatta acccaaaattt gcataacccc 180
atgtt当地tttt tc当地tcaacca cctctt当地ctt gt当地gatttct gctattt当地tctc catcattt当地gtct 240
cccatgatgc tggta当地acct ggtt当地tagaa gatagaacca tttc当地tatttctc aggatgtt当地ttg 300
gtgcaatttct tttt当地ttttg cacctt当地gtta gt当地actgaat taattt当地ttatt tgccgt当地gtatg 360
gccttatgacc actttgtggc cattt当地gaat cctctt当地gtctc acacagtttgc catctccc当地ag 420
aaactctgtg ccatgctt当地gg ggtt当地tattt gatgcat当地ggg gagtc当地gc当地atg tt当地ccctt当地gaca 480
ctcgctgtct ct当地gtt当地aaa gttat当地ttttt catggtt当地ca acacaat当地caa tc当地at当地tttctc 540
tgtgaggat tatcactt当地ctc taccctt当地gact ct当地tat当地tc当地ag ccagttt当地gtctt 600
ctttt当地caactg tt当地gccacttta taatgagata accacacttac tc当地atc当地tatttctt gacatctt当地tat 660
gcattt当地catca tt当地gtc当地accac cttgaagatg cttc当地agcca gt当地gggaccacg caaaat当地gttctc 720
tccacctt当地gtg cctt当地ccactt gactgctt当地atc accatctt当地cc att当地ggc当地accat cctt当地ctt当地ctc 780
tactgtt当地tac ccaactt当地ccaa aaactt当地ccagg cacacagttca aagt当地ggc当地ctc tttt当地gtt当地ttac 840
accgtt当地gtga tccctt当地gtttt gaatccctt gatc tt当地tacagttca tgagaaaataa agatgtt当地aaag 900
gatgcaatcc gaaaaataaataa caatacaaaa tattttt当地cata ttaaaatcatag gcattt当地ggat 960
ccat当地tttaattt tt当地gtt当地atttga acaataa 987

```

<210> 45
<211> 975
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 7476079CB1

<210> 46

<211> 948
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476112CB1

<400> 46
atgcaggggc taaaccacac ctccgtgtct gaattcatcc tcgttggctt ctctgccttc 60
ccccacctcc agctgatgtcttccctgtttccctgctga tgtaacctt caccgtgtc 120
ggcaacctgc tcatcatggc cactgtctgg agcgagcgca gcctccacat gcccatgtac 180
ctctccctgt gtgcctctc catcaccggatccctctaca ccgtggccat catcccgcgc 240
atgtggcccg acctgctgtc cacceagcgc tccatcgcc ttctggcctg tgccagttag 300
atgttcttct ctttcagctt cggcttcacc cactccttcc tgctcaactgt catgggctac 360
gaccgctacg tggccatctg ccacccctg cgttacaacg tgctcatgag cctgcggggc 420
tgccacctgcc gggggggctg ctccctggctt ggtggcttgg tcatggggat ggtgggtgacc 480
tcggccatttttccacctcgc cttctgtggaca aaggaga tccaccattt ctctgccac 540
gtgcacccctc tggtgaagg ggccctgtggatgatgtgtcc tggtggggc caaaggcggtg 600
ggcttgggtgt gtatcacggc cctgctggc tgtttctcc tcatcccttctccatgcc 660
ttcatcggtggccgcattttaaagatccct tctgctgtggatgatgtgtcc tggtggggc 720
acctgtgcctt ctcacccctac tgggtgggtc gtgcactatg gctttgcctc cgtcattttac 780
ctgaagccca aagggtccca gtctccggaa ggagacaccc ttatgggcat cacctacacg 840
gtccotcacac ctttcctcag ccccatcatc ttccgcctca ggaacaagga gctgaagggtc 900
ggcatgaaga agacttgctt caccaaactc ttccacaga actgctga 948